

Intelligence and its Assessment Across the Life Span:  
Analyzing Construct and Criterion Validity

**Inaugural Dissertation**

Submitted to  
the Faculty of Psychology at the University of Basel  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy

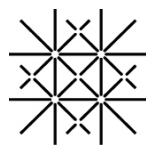
by

Jasmin Tamara Gygi

from Kappelen, BE

Basel, Switzerland

2016



**University  
of Basel**

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel [edoc.unibas.ch](http://edoc.unibas.ch)

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0

Approved by the Faculty of Psychology  
at the request of

Prof. Dr. Rui Mata (Chair)

Prof. Dr. Alexander Grob (Referee)

Prof. Dr. Sakari Lemola (Co-Referee)

Basel, 31.01.2017

---

Prof. Dr. Roselind Lieb (Dean)

## ACKNOWLEDGMENTS

With lots of gratitude to the following people who supported me throughout my PhD:

Prof. Dr. Alexander Grob—for your guidance and continuous support and the effort you put into my research, and for being my doctoral advisor.

Prof. Dr. Sakari Lemola—for your encouraging words, for the inspiring academic discussions and input, and for acting as co-referee of this dissertation.

Prof. Dr. Rui Mata—for serving on the dissertation committee.

Dr. Priska Hagmann-von Arx—for being my mentor all along, for your everlasting encouragement and supportive feedback, and for giving me confidence in my work.

Dr. Olivia Manicolo—for inspiring and encouraging me in every sense as well as for your valuable comments when proofreading my dissertation.

My project team—for making the project a success with your endless dedication.

My colleagues—for the academic exchange and the creative environment. Especially

Dr. Rebekka Weidmann for your support and motivation.

My mother, Marie-Thérèse Gygi—for your emotional support in every time and place, for always believing in my potential, and for giving me the strength I needed to finish this work. To you I dedicate this dissertation.

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	III
ABSTRACT .....	VI
1. Introduction .....	1
2. Theoretical Background .....	4
2.1 Intelligence .....	4
2.2 Intelligence Assessment .....	7
2.3 Validity of Intelligence Assessment .....	11
2.3.1 Construct Validity With a Focus on Factorial Validity .....	12
2.3.2 Criterion Validity With a Focus on Concurrent and Predictive Validity. ....	14
3. Research Questions .....	19
4. Method .....	21
4.1 Studies and Samples .....	21
4.3 Measures .....	22
4.4 Statistical Analyses .....	25
5. Synopsis of Results .....	26
5.1 Factorial Validity of the RIAS .....	26
5.2 Concurrent and Predictive Validity of Currently Used Intelligence Tests .....	27
6. General Discussion .....	28
6.1 Factorial Validity of the RIAS .....	28
6.2 Concurrent and Predictive Validity of Currently Used Intelligence Tests .....	30
6.3 Strengths and Limitations .....	35
6.4 Conclusion and Outlook .....	38
References .....	40
APPENDIX A: Study 1 .....	56
APPENDIX B: Study 2 .....	77

APPENDIX C: Study 3 .....	91
APPENDIX D: Study 4 .....	107
APPENDIX E: Selbständigkeitserklärung .....	117

## ABSTRACT

Intelligence is the strongest predictor of a diverse range of positive outcomes in life, such as scholastic achievement, career success, health, and longevity. Hence, many conclusions are drawn from intelligence test results and these can have an impact on high-stakes decisions regarding, for example, a child's school career or an adult's employment. To accurately interpret the results of an intelligence test, its validity has to be ensured. This cumulative dissertation includes four studies that extend current knowledge on the construct and criterion validity of currently used intelligence tests: The factor structure of the Reynolds Intellectual Assessment Scales (RIAS) and its measurement invariance were evaluated across its four language versions (Danish, English, German, and Spanish) and across individuals with and without a migration background in the German RIAS version. Further, the predictive power of four intelligence tests currently used in German-speaking countries was analyzed for longitudinal scholastic achievement (i.e., school grades), as well as the predictive power of the RIAS and the incremental validity of conscientiousness for career success (i.e., occupational status, income, job satisfaction) in adults. Regarding construct validity, findings support the RIAS factor structure as well as measurement invariance across its four language versions and across German-speaking individuals with and without a migration background. Further, individuals with a migration background showed lower verbal and nonverbal intelligence. Regarding criterion validity, currently used intelligence tests were a predictor for longitudinal school grades. Furthermore, the RIAS showed positive associations with concurrent occupational status, while conscientiousness could explain variance in income and job satisfaction. This points to further factors that might be relevant for psychological assessment beyond intelligence. In conclusion, the present dissertation provides evidence for construct and criterion validity of currently used intelligence tests in German-speaking countries, indicating a valid intelligence assessment across the life span.

## 1. Introduction

Intelligence is one of the constructs most often studied in psychology and most often measured by practitioners (Goldstein, Princiotta, & Naglieri, 2015). Intelligence is described as the ability to solve problems and adapt to new situations by reasoning, learning from experience, and planning ahead, and is thus more than merely possessing academic knowledge (Gottfredson, 1997). Hence, intelligence is a latent construct and can be assessed through intelligence tests that create a situation in which intelligent behavior can be observed. Since the first intelligence test was developed for children by Binet and Simon (1905), the primary purpose of intelligence assessment has been to predict scholastic and academic achievement to foster the optimal development of each individual and to determine the best school setting for that person (Binet & Simon, 1905). Currently, there is evidence that intelligence is a predictor of important life outcomes: Intelligence has been found to correlate positively with academic achievement, career success, physical fitness, and health, and negatively with obesity, drug addiction, and mortality (Batty, Deary, & Gottfredson, 2007; Deary, 2009; Gottfredson & Deary, 2004). However, to draw accurate conclusions based on intelligence test results, test procedures have to provide valid intelligence test scores. Thus, test scores and their interpretation must be consistent across individuals, populations, or contexts (Messick, 1995). In this vein, intelligence tests are often validated in terms of content, construct, and criterion validity (Braden & Niebling, 2005). Content validity refers to the representativeness of a test's items and can be analyzed on the basis of theory or expert opinion. Construct validity refers to the interpretation of test scores in accordance with a test's theoretical structure. Evidence of a test's construct validity can be supported, for example, through confirmatory factor analysis. Criterion validity is given when test scores can predict a criterion related to a construct outside the testing situation, such as the prediction of school grades or occupational status based on intelligence test results (Moosbrugger & Kelava, 2012).

Currently, a multitude of intelligence tests are available (e.g., Hagmann-von Arx, Gauck, & Grob, 2015), some specifically designed for children, adolescents, or adults, and others for all age groups across the life span. The Reynolds Intellectual Assessment Scales (RIAS; Reynolds & Kamphaus, 2003) is a time-efficient intelligence test that measures general intelligence as well as fluid and crystallized intelligence across the life span. In addition, memory can be assessed independently from intelligence. The RIAS has been adapted to Danish (Hartmann & Andresen, 2011), German (Hagmann-von Arx & Grob, 2014), and Spanish (Santamaría & Fernández Pinto, 2009). However, to date, no study has examined the construct validity across the test's four currently available language versions or across individuals with and without a migration background. Yet as migration continues to rise, it is increasingly important to understand how migration background affects intelligence test scores to ensure accurate interpretations of test results.

The following are also well known and often used intelligence tests in German-speaking countries: The Intelligence and Development Scales (IDS; Grob, Meyer, & Hagmann-von Arx, 2013), the Snijders-Oomen Nonverbal Intelligence Test Revised 6–40 (SON-R 6–40; Tellegen, Laros, & Petermann, 2012), and the Wechsler Intelligence Scales for Children, Fourth Edition (WISC-IV; Wechsler, 2003). To the best of my knowledge, there is a paucity of literature on the predictive power of intelligence tests currently used in German-speaking countries for scholastic achievement (i.e., school grades) and career success (i.e., occupational status, income, job satisfaction).

This dissertation expands upon current knowledge on intelligence assessment by examining construct and criterion validity of intelligence tests currently used in psychological practice. Regarding construct validity (i.e., factorial validity), Study 1 examined the structure of the RIAS and its measurement invariance across different language groups (Gygi, Ledermann, Grob, & Hagmann-von Arx, *submitted*; see Appendix A). Study 2 examined the structure of the RIAS, its measurement invariance, and latent mean differences across



individuals with and without a migration background (Gygi, Fux, Grob, & Hagmann-von Arx, 2016; see Appendix B). Regarding criterion validity, Study 3 analyzed predictive validity of several intelligence tests (i.e., IDS, RIAS, SON-R 6–40, WISC-IV) for scholastic achievement (Gygi, Hagmann-von Arx, Schweizer, & Grob, *submitted*; see Appendix C), and Study 4 analyzed concurrent validity of the RIAS for career success (Hagmann-von Arx, Gygi, Weidmann, & Grob, 2016; see Appendix D).

Chapter 2 summarizes the relevant theoretical background and Chapter 3 presents the research questions that motivated this dissertation. Chapter 4 gives an overview of the studies and samples, and describes measures and statistical analyses used. Chapter 5 is a synopsis of results, and Chapter 6 concludes with a general discussion of the main findings, their practical implications, and an outlook for future research.

## 2. Theoretical Background

The sections in this chapter summarize the theoretical background relevant to this dissertation: First the construct of intelligence will be outlined and then its assessment will be reviewed. Finally, the chapter concludes with the validation of intelligence assessment.

### 2.1 Intelligence

Although intelligence has been well studied and frequently measured (Goldstein et al., 2015), still there is no holistic and entirely accepted definition of intelligence. However, in 1997 Gottfredson republished a 1994 *Wall Street Journal* editorial that was signed by 52 researchers in which mainstream ideas on intelligence were outlined. This editorial provided a basic definition that has been widely cited since:

Intelligence can be defined as a very general mental capacity that among other things involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill or test taking smarts. Rather it reflects a broader and deeper capability for comprehending our surroundings, catching on, making sense of things, or figuring out what to do. (Gottfredson, 1997, p. 20)

In sum, intelligence is an individual's ability to adapt to new situations by reasoning, learning from experiences, and planning ahead and comprises not only academic knowledge.

Theories of intelligence and its structure date back to Spearman (1904), who concluded that intellectual abilities can be represented by a general intelligence factor, *g*. Originally, Spearman's theory was called the *two-factor theory*, as it separated variance of intelligence into a general intelligence factor, *g* (shared variance across measures) and specific factors, *s* (variance unique to measures; Wasserman & Tulsy, 2005). Over time, more theories emerged, were further developed, and consequently combined. Influential theories are the *Cattell–Horn gf-gc theory* (Horn & Cattell, 1966; Horn & Noll, 1997),

*Carroll's three-stratum theory* (Carroll, 1993), and the *Cattell–Horn–Carroll theory* (CHC; McGrew, 2005). Horn and Cattell (1966) postulated the gf-gc theory. They disagreed about an overall general intelligence factor and divided g into fluid and crystallized intelligence (McGrew, 2005). Fluid intelligence encompasses problem solving and reasoning and is barely influenced by acquired knowledge, being more dependent on heredity. The development of crystallized intelligence, in turn, is first influenced by fluid intelligence and represents acquired knowledge of language, culture, and concepts. Consequently, it is also strongly influenced by education and culture (Cattell, 1987). Horn (1991) extended the gf-gc theory and proposed not only fluid and crystallized intelligence, but also 10 primary abilities, which comprise numerous narrow abilities. Carroll (1993) posited the three-stratum theory that categorized intellectual abilities in three strata: Stratum I encompasses numerous narrow abilities; Stratum II encompasses nine broad factors, including similar factors to those defined by Horn (1991); Stratum III comprises the general intelligence factor, g, as defined by Spearman (1904). Today's most significant theory of intelligence is the CHC theory (McGrew, 2005) that combines the Cattell–Horn gf-gc theory with Carroll's three-stratum theory in one comprehensive theory. The CHC theory categorizes intelligence in the three strata laid out by Carroll (1993): There are up to 81 narrow abilities on Stratum I (e.g., *language development*, *perceptual speed*, *memory span*, and *induction*); 10 broad abilities on Stratum II (i.e., *fluid intelligence* [Gf], *comprehension–knowledge* [Gc], *short-term memory* [Gsm], *visual processing* [Gv], *auditory processing* [Ga], *cognitive processing speed* [Gs], *decision and reaction speed* [Gt], *reading and writing*, [Grw], and *quantitative knowledge* [Gq]); and g, as defined by Spearman (1904), on Stratum III. Later, Schneider and McGrew (2012) extended the CHC theory by six additional broad abilities on Stratum II, which, however, have yet to be further researched (i.e., *domain-specific knowledge* [Gkn], *olfactory ability* [Go], *tactile ability* [Gh], *psychomotor ability* [Gp], *kinesthetic ability* [Gk], and *psychomotor speed* [Gps]).

Alongside the development of different theories of intelligence, multiple intelligence tests emerged. Many of them assess general intelligence following Spearman (1904). The more recently developed intelligence tests have been influenced primarily by the CHC theory and thus also assess some of the broad abilities (Willis, Dumont, & Kaufman, 2011). These include, for example, the RIAS (Reynolds & Kamphaus, 2003), the Stanford–Binet Intelligence Scales, Fifth Edition (SB5; Roid, 2003a), and the WISC-IV (Wechsler, 2003).

Intelligence is considered one of the most stable traits in people (Goldstein et al., 2015). Two types of stability can be distinguished: Mean-level stability and rank-order stability (Rost, 2009). Mean-level stability refers to the development of intelligence: Cognitive abilities increase rapidly in the first years of life until late adolescence and then reach a plateau. With advancing age, fluid intelligence declines, while crystallized intelligence remains stable and does not decline until old age (Deary, 2014; Rost, 2009). Rank-order stability, in turn, has to do with interindividual differences and whether individuals remain stable in their relative order to each other as they advance in age. Stability in an individual's rank order depends strongly on a test's quality, on the interval between two measurements, and on the age at time of the first measurement (Deary, 2014). It has been shown that rank order of intelligence is not very stable for toddlers until the age of 4 years, first because intelligence is assumed to be rather unstable in young age and second because the currently available tests for young infants are lacking in terms of quality. With the beginning of primary school, stability of intelligence increases (Rost, 2009). For example, Moffit, Caspi, Harkness, and Silva (1993) found high stability ( $r = .74$  to  $.85$ ) in intelligence for children aged 7 to 13 years. Also for older ages, a study found strong correlations over several decades of  $r = .54$  to  $.67$  between scores of the same test assessed at ages 11, 70, and 90 years, respectively (Deary, Pattie, & Starr, 2013; Deary, 2014).

In sum, intelligence can be described as the ability to adapt to new situations and it encompasses reasoning and acquired knowledge. The structure of intelligence can be depicted

as being one general intelligence factor on the one hand and as additionally including many specific intelligence factors on the other, which differ depending on the respective theory. Further, intelligence has shown to be a very stable construct from primary school age on and thus has been postulated to be a significant predictor of multiple life outcomes, as is further delineated in the next sections.

## 2.2 Intelligence Assessment

It can be inferred from the above-mentioned theories (i.e., Spearman's two-factor theory, the Cattell–Horn gf-gc theory, Carroll's three-stratum theory, the CHC theory) that intelligence is a latent construct and thus, not directly observable. Therefore, intelligence tests are constructed to provide situations in which *intelligent behavior* can be observed (Schneider & Flanagan, 2015). Hence, intelligence tests provide scores (IQ scores) that are drawn from the composite intelligence index of a test and are seen as an estimate of *g*. IQ scores are normally distributed in the population and are usually standardized with  $M = 100$  and  $SD = 15$ . Consequently, 68% of the population lies within one standard deviation below and above the mean (IQ = 85 to 115) and 95% of the population lies within two standard deviations below and above the mean (IQ = 70 to 130). IQ scores below 70 are considered mental retardation, whereas scores above 130 are considered mental giftedness, each represented by approximately 2% of the population (Rost, 2009).

The first intelligence test was developed in France by Binet and Simon (1905) with the intention to detect children who could profit from special education. Though the *Binet–Simon Scale* was innovative and efficient, Terman of Stanford University saw the need for improvements and consequently translated, revised, and published the *Stanford Revision and Extension of the Binet–Simon Intelligence Scale* in 1916 in the United States. Although many translations and further developments of the Binet–Simon Scale were carried out contemporaneously (Goldstein et al., 2015), Terman's (1916) Stanford–Binet was number one

for intellectual assessment and was revised several times in the following decades (Roid, 2003b): The *Revised Stanford–Binet Intelligence Scale* (Terman & Merrill, 1937), the *Stanford–Binet Intelligence Scale, Form L–M* (Terman & Merrill, 1960), and the *Stanford–Binet Intelligence Scale: Fourth Edition* (SB-IV; Thorndike, Hagen, & Sattler, 1986). The most recent, fifth edition (SB5), was published by Roid (2003a). The SB5 is an individually administered intelligence test for individuals aged 2 to above 85 years and thus is appropriate for intelligence assessment across the life span. Further, the SB5 is constructed according to the CHC theory and contains five factors that are measured nonverbally and verbally, respectively: *fluid reasoning*, *knowledge*, *quantitative reasoning*, *visual–spatial processing*, and *working memory*. These five broad abilities are on Stratum II, their specific subtests on Stratum I, and general intelligence on Stratum III. The SB5 contains many features of its previous editions and also some significant enhancements, such as its level design, the routing subtests, child-friendly material, and a wide breadth of scale. It has been shown that the SB5 is highly sensitive in the lower and higher scale ranges and thus is especially applicable for assessing the full intelligence scale range from mental retardation to intellectual giftedness (Mleko & Burns, 2005). Currently, the SB5 is being translated, adapted, validated, and standardized for the German-speaking part of Switzerland and Germany at the University of Basel, Switzerland (Grob, Gygi, & Hagmann-von Arx, in preparation). In addition to studying for my Ph.D., I have been responsible for the entire undertaking regarding the SB5, under the direction of Prof. Dr. Alexander Grob and in close cooperation with Dr. Priska Hagmann-von Arx. The translation and adaptations started in 2013, followed by several pretests and a pilot study in 2014. After changing culturally biased items and reordering items by their difficulty—according to the results of the pilot study—the standardization ( $N_{\text{as of December 2016}} = 763$ ) and validation ( $N_{\text{as of December 2016}} = 203$ ) started in 2015.

Besides being revised by Terman, the Binet–Simon Scale was further developed for the selection and recruitment of World War I soldiers, resulting in group-administered tests

called *Army Alpha*, containing verbal content, and *Army Beta*, containing nonverbal content for immigrants (Yoakum & Yerkes, 1920; Flanagan & Kaufman, 2004). Giving weight to both the Army Alpha and Army Beta tests, David Wechsler created the *Wechsler Bellevue Scales* (Wechsler, 1939). A series of intelligence tests for individuals across the life span followed: The current editions are the *Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition* for preschoolers aged 2 to 7 years (WPPSI-IV; Wechsler, 2012), the *Wechsler Intelligence Scale for Children, Fifth Edition* for children aged 6 to 16 years (WISC-V; Wechsler, 2014), and the *Wechsler Adult Intelligence Scale, Fourth Edition* for adults aged 16 to 90 years (WAIS-IV; Wechsler, 2008). In the following, the focus is on the WISCs, as in this dissertation the remaining scales were not applied. The Wechsler Intelligence Scales in their different editions have been translated and adapted worldwide and currently exist in many languages (Flanagan & Kaufman, 2004), such as the German version of the WISC-IV (Petermann & Petermann, 2013). The WISC-IV assesses general intelligence on Stratum III, four broad abilities on Stratum II (i.e., *verbal comprehension, perceptual reasoning, working memory, processing speed*), and their specific subtests on Stratum I, following the CHC theory. The advantages are its theory-driven development and the extraction of four specific intelligence indices besides the composite intelligence index, thus allowing the analyses of intelligence profiles (Hagmann-von Arx et al., 2015). Such profile analyses, however, are less reliable as *g* (Borsuk, Watkins & Canivez, 2006; Watkins, 2000).

Another currently used intelligence test applicable across the life span is the RIAS (Reynolds & Kamphaus, 2003). It is standardized for individuals aged 3 to 94 years in the United States and was recently adapted and standardized in Denmark (Hartmann & Andresen, 2011), Switzerland and Germany (Hagmann-von Arx & Grob, 2014), and Spain (Santamaria & Fernández Pinto, 2009). The RIAS was constructed according to the CHC theory as well and therefore consists of four intelligence subtests on Stratum I (i.e., *guess what, verbal reasoning, odd-item out, what's missing*), general intelligence on Stratum III, and its two

components, verbal and nonverbal intelligence, on Stratum II, which are based on crystallized and fluid intelligence, respectively. Furthermore, the RIAS measures memory independent of *g* through two additional subtests. Thus, for RIAS intelligence, two structures can be summarized: (a) a single-factor structure with the four intelligence subtests each loading on a general intelligence factor; and (b) a two-factor structure with the two verbal intelligence subtests each loading on a verbal intelligence factor and the two nonverbal intelligence subtests each loading on a nonverbal intelligence factor. Advantages of the RIAS are its time-efficient administration, its user friendliness, and its independence of visual-motor speed and reading ability (Andrews, 2007).

An intelligence test developed at the University of Basel, Switzerland, and standardized in German-speaking countries is the IDS for children aged 5 to 10 years (Grob et al., 2009, 2013). The IDS for preschool (IDS-P) was developed for children aged 3 to 5 years (Grob, Reimann, Gut, & Frischknecht, 2013). The IDS embodies an entire redevelopment of the Kramer Tests (Kramer, 1972), which in turn represented further developments of the Binet–Simon Scale (1905) for German-speaking countries. The IDS assesses general intelligence, primarily fluid intelligence, with seven subtests (i.e., *visual perception*, *selective attention*, *phonological memory*, *visual-spatial memory*, *auditory memory*, *abstract reasoning*, *figural reasoning*) following Spearman’s (1904) two-factor theory. In addition to intelligence, the IDS assesses five developmental domains including *psychomotor skills*, *social–emotional competence*, *mathematics*, *language*, and *motivation*. Benefits of the IDS is the holistic approach to assessing intelligence as well as other important developmental domains during childhood, allowing the analyses of developmental profiles (Grob et al., 2009, 2013). Currently, revision and further development of the IDS are taking place at the University of Basel, Switzerland. The *Intelligence and Development Scales for Children and Adolescents* (IDS-2; Grob & Hagmann-von Arx, in preparation) have been expanded for individuals aged 5 to 20 years and are currently being standardized and validated. Moreover,



the IDS-2 is being translated and adapted in other European countries, such as Finland, Sweden, and Poland.

Finally, another series of intelligence tests developed in Europe and standardized in the Netherlands and Germany is the *Snijders-Oomen Nonverbal* (SON) tests. The first edition of the SON (Snijders-Oomen, 1943) was originally developed for deaf children aged 4 to 14 years, as intelligence can be assessed nonverbally. Several revisions followed, resulting in two currently available editions appropriate for individuals with and without deafness: The revised SON for children aged 2 to 7 years (SON-R 2½–7; Tellegen, Laros, & Petermann, 2007) and the revised SON for individuals aged 6 to 40 years (SON-R 6–40, Tellegen et al., 2012). The SON-R 6–40 consists of four subtests (i.e., *analogies*, *mosaic*, *categories*, *draw pattern*) that primarily measure fluid intelligence following the Cattell–Horn gf-gc theory (Horn & Cattell, 1966), although its development did not follow any specific theory (Petermann & Renner, 2010). The main advantages of the SON-R tests are clearly their time-efficient and nonverbal administration. Thus, their use is especially favorable when assessing individuals with hearing and language deficiencies or with German as a foreign language (Hagmann-von Arx et al., 2015).

### **2.3 Validity of Intelligence Assessment**

To draw accurate conclusions based on intelligence test results, a test's validity has to be ensured. Messick defined validity as “an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment” (1995, p. 1). Thus, test scores must be consistent across individuals, populations, or contexts (Messick, 1995). Different forms of validity that are frequently used to validate psychological test procedures such as intelligence tests include content, construct, and criterion validity (Braden & Niebling, 2005). Content validity captures the representativeness of a test's items and can

be assessed, for example, on the basis of theory or experts' opinions. The process of ensuring content validity is usually theoretical in nature and thus not analyzed through empirical studies but often throughout a test's development (Moosbrugger & Kelava, 2012). However, after test construction, independent empirical studies are needed to support construct and criterion validity. The next two subsections present construct and criterion validity in more detail, illustrate the current literature, and identify possible gaps in research.

**2.3.1 Construct Validity With a Focus on Factorial Validity.** Construct validity implies that test scores can be reliably interpreted along with the test's theoretical structure. Construct validity is supported if tests measuring the same construct are highly correlated (i.e., convergent validity)—or if tests measuring distinct constructs are not or only weakly correlated (i.e., divergent validity; Moosbrugger & Kelava, 2012). For example, Hagmann-von Arx, Lemola, and Grob (2016) analyzed five currently used intelligence tests in German-speaking countries (i.e., IDS, RIAS, SON-R 6–40, WISC-IV, Culture-Fair Intelligence Test Scale 2) in children aged 6 to 11 years. They found strongly correlated IQ scores ( $r = .77$  to  $.90$ ), thus supporting convergent validity of these five intelligence tests.

A test's construct validity can also be analyzed through exploratory and confirmatory factor analyses. Exploratory factor analyses are conducted to reveal underlying structures and relationships between variables. Confirmatory (but not exploratory) factor analyses are consulted to ensure factorial validity of a chosen structure (Moosbrugger & Kelava, 2012). Structures that resulted as reliable from such factor analyses can be further analyzed for their measurement invariance across different groups of individuals. When measurement invariance is found, one can be reassured that group differences in observed means are due to mean differences in their latent construct, allowing a comparison of these groups (Widaman & Reise, 1997). For example, validation of the two factor structures of the English version of the RIAS can be found in independent studies and not only in the test's technical manual

(Reynolds & Kamphaus, 2003): The two-factor structure was supported by a study using a clinical sample (Beaujean, McGlaughlin, & Margulies, 2009). Several studies found evidence supporting the single-factor structure using the RIAS standardization sample (Dombrowski, Watkins, & Brogan, 2009), clinical samples (Beaujean & McGlaughlin, 2014; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007), or samples with typically developing individuals (Irwin, Joschko, & Kerns, 2014). However, the other three RIAS language versions have not yet been evaluated for their underlying factor structure other than in the technical manuals (Hagmann-von Arx & Grob, 2014; Hartmann & Andresen, 2011; Santamaría & Fernández Pinto, 2009), nor has the invariance across the four currently available RIAS language versions been examined. Also, the RIAS factor structure and its measurement invariance across individuals with and without a migration background has not yet been studied. However, migration is rising in recent decades in Europe (Eurostat, 2014). For example, in Switzerland and Germany, 25% and 21% of the population, respectively, have a migration background (Bundesamt für Statistik, 2015; Statistisches Bundesamt, 2015). This leads to more individuals that are seeking psychological assessment in a foreign-language environment. Language differences, in turn, can lead to difficulties during assessment due to inferior abilities in verbal responses and misunderstanding of task instructions. Hence, individuals with a migration background may not be able to show their full potential, what may lead to test scores that underestimate their intelligence and may consequently cause missed educational opportunities (Calero, Fernández-Parra, et al., 2013; Georgas, Van De Vijver, Weiss, & Saklofske, 2003; Hagmann-von Arx, Petermann, & Grob, 2013; Weiss et al., 2006). Lower test scores in individuals with a migration background are especially apparent in tasks with high language requirements, while their test performance increases with decreasing language requirements of the tasks (Daseking, Lipsius, Petermann, & Waldmann, 2008; Hagmann-von Arx et al., 2013; Harris, Muñoz, & Llorente, 2008). Also cultural and environmental factors may negatively affect test performance of individuals with

a migration background (Mackintosh, 2011). Individuals with a migration background may, for example, be less acquainted with these type of tasks used in performance assessment (Calero, Mata, et al., 2013; Resing, Tunteler, de Jong, & Bosma, 2009). Further, immigrant groups are confronted with negative achievement stereotypes, what may lead to increased performance pressure (Appel, Weber, & Kronberger, 2015). Parental language skills have also been linked to decreased performance: Parents who have limited language skills in the local language have shown to be less involved in schooling, what may again adversely affect their children's performance (Lahaie, 2008; Turney & Kao, 2009). However, high-stakes decisions may derive from intelligence test scores. For example, underestimated intelligence test scores may result in inaccurate schooling placements and missed educational opportunities (Hessels, 1997; Klingner, Blanchett, Harry, 2007) or in unjustified rejections of job applications (Baltes & Rudolph, 2010). Understanding how migration status influences test scores is therefore crucial to correctly interpreting the results.

A first goal of this dissertation was to fill these gaps: The two factor structures of the RIAS were examined across (a) the four RIAS language versions and (b) individuals with and without a migration background in the German RIAS. Subsequently, measurement invariance across these different groups was analyzed to ensure factorial validity of the RIAS as well as the comparability of RIAS test results in individuals with and without a migration background.

**2.3.2 Criterion Validity With a Focus on Concurrent and Predictive Validity.** Criterion validity is given when test scores can predict a criterion that is relevant for diagnostic decisions outside the testing situation, such as, the prediction of school grades by an intelligence test's scores. In more detail, if the intelligence test scores and the school grades were assessed at the same time point, it is called concurrent validity. If the school grades were assessed later in time, it is called predictive validity (Moosbrugger & Kelava, 2012).

Intelligence has been shown to be one of the most important predictors for a wide range of life outcomes (Deary, 2012; Gottfredson & Saklofske, 2009). For example, intelligence has been found to be a strong positive predictor of scholastic and work achievement, physical fitness, and health and a strong negative predictor of mortality, obesity, and drug addiction (see Batty et al., 2007; Deary, 2009; Gottfredson & Deary, 2004). Studies analyzing the concurrent and predictive validity of intelligence for scholastic achievement and career success most often analyze *g*, or a test's composite intelligence index, as described in the CHC theory (see Section 2.2).

Regarding scholastic achievement, studies have shown moderate to strong correlations (around  $r = .40$  to  $.70$ ) between intelligence and scholastic and academic achievement (e.g., Mackintosh, 2011; Sternberg, Grigorenko, & Bundy, 2001). Studies have found not only cross-sectional but also longitudinal correlations of intelligence and academic achievement: In a study over 5 years, cognitive ability strongly correlated ( $r = .81$ ) with overall educational achievement (Deary, Strand, Smith, & Fernandes, 2007). Further, meta-analytical results showed strong correlations for primary education ( $r = .56$  to  $.58$ ) but only weak correlations for secondary and tertiary education ( $r = .23$  to  $.24$ ; Poropat, 2009; Strenze, 2007). These results can be explained by the decreasing variance in IQ scores, as predominantly students with higher intelligence pursue higher education (Goldstein et al., 2015; Mackintosh, 2011). Thus, with restriction of range, correlations between intelligence and achievement may weaken (Sternberg et al., 2001). Another, more recent meta-analysis (Roth, Becker, Romeyke, Schäfer, Domnick, and Spinath, 2015) found a moderate observed correlation ( $r = .44$ ) between intelligence and school grades. However, in contrast to previous studies (Poropat, 2009; Strenze, 2007), results of Roth et al. (2015), revealed that correlations between intelligence and school grades tend to be weaker in primary school ( $r = .40$ ) compared to middle and high school ( $r = .46$ ) students. They suggest that intelligence deficits in primary school students may be compensated for more easily through practice than in higher grade-

level students, because of less complex learning content. Furthermore, regarding subject domains, their results support highest and comparable correlations for mathematics/science ( $r = .42$ ) and languages ( $r = .36$ ).

Studies have also shown differences in the strength of correlations depending on the operationalization of scholastic achievement: Correlations between intelligence and scholastic achievement tend to be higher when standardized achievement tests are analyzed instead of school grades (Rost, 2009; Sternberg et al., 2001). Standardized achievement tests represent achievement at only one point in time, while school grades represent achievement based on learning over longer periods of time. Nevertheless, school grades are crucial for a child's academic promotion each year as well as for further scholastic and occupational opportunities (Roth et al., 2015).

However, research on currently used intelligence tests in German-speaking countries is scarce: For the IDS, two studies have supported concurrent ( $\beta = .30$  to  $.56$ ) and predictive validity over 3 years ( $\beta = .21$  to  $.34$ ) of intelligence for scholastic achievement assessed in school-aged children through parental ratings and school grades averaged across mathematics, science, and language (Gut, Reimann, & Grob, 2012, 2013). These results indicate small to moderate predictive validity of the IDS general intelligence for averaged school grades.

The technical manual of the English version of the RIAS (Reynolds & Kamphaus) reports strong correlations between the composite intelligence index and a standardized achievement test in mathematics ( $r = .67$ ) and language ( $r = .64$ ) for school-aged children. This suggests strong predictive validity of the RIAS composite intelligence index for standardized achievement tests.

For the SON-R 6–40, the technical manual of the German version (Tellegen et al., 2012) reports moderate to strong correlations between test scores and concurrent school grades in mathematics ( $r = .58$ ) and language ( $r = .49$ ) for primary school children. Thus,

nonverbal intelligence measured based on the SON-R 6–40 moderately to strongly predict school grades.

For the English WISC-IV, concurrent validity of the general intelligence index and its specific indices on a standardized academic achievement test in mathematics and reading has been shown (Glutting, Watkins, Konold, & McDermott, 2006). General intelligence explained 60% of the variance in scholastic achievement, while the specific indices explained only 0 to 2% additional variance. To the best of my knowledge, no independent studies on predictive validity for school grades have been made for the German versions of the RIAS, SON-R 6–40, and WISC-IV. Hence, another aim of this dissertation was to extend knowledge on the criterion validity of currently used intelligence tests: Predictive validity of the German versions of the IDS, RIAS, SON-R 6–40, and WISC-IV on school grades was analyzed for German-speaking countries.

In adult samples, intelligence is associated with achievement in the working field such as career success. Career success can be distinguished as objective and subjective achievement: An objective achievement is, for example, occupational skill level (occupational status) or income, whereas subjective achievement is represented by, for example, an individual's job or career satisfaction (Ng, Eby, Sorensen, & Feldman, 2005). Schmidt and Hunter (2004) found that intelligence is a strong predictor of occupational skill level and reported meta-analytical results for both concurrent and predictive validity of around  $r = .50$ . The meta-analysis by Ng et al. (2005) found a small to medium correlation between intelligence and income of  $r = .27$ . Another meta-analysis showed similar results with correlations of  $r = .45$  for occupational skill level and  $r = .23$  for income (Strenze, 2007). The association of intelligence with job satisfaction is moderated by job complexity and occupational skill level: Studies have indicated that intelligence is negatively correlated with satisfaction in jobs with low complexity and a low occupational skill level ( $r = -.30$ ), whereas it is positively correlated in jobs with high complexity and a high occupational skill level ( $r =$

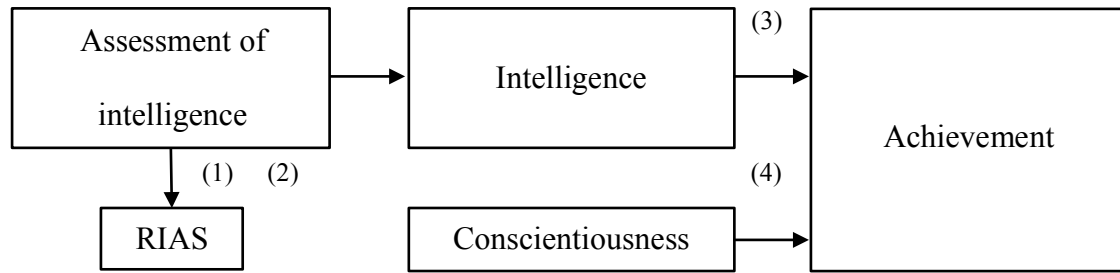
.30) (Ganzach & Fried, 2012; Lounsbury, Gibson, Steel, Sundstrom, & Loveland, 2004). This indicates that, for example, people with a higher intelligence show lower job satisfaction when holding a job with low complexity, while they show higher job satisfaction when holding a job with high complexity. Regarding specific factors of intelligence, a recent theoretical model by Schmidt (2014) proposed that primarily crystallized intelligence has a direct influence on achievement such as career success, but there is a lack of studies analyzing fluid and crystallized intelligence separately. Beyond intelligence, other traits have also been shown to predict career success, such as personality (Poropat, 2009). From the Big Five model of personality (Costa & McCrae, 1995) in particular conscientiousness is correlated with career success, such that more conscientious people report higher career success, including occupational skill level, income, and job satisfaction (about  $r = .20$ ,  $.07$ , and  $.26$ , respectively; Hurtz & Donovan, 2000; Ng et al., 2005; Sackett & Walmsley, 2014). Conscientiousness can be further divided into its six narrow facets (i.e., competence, order, dutifulness, achievement striving, self-discipline, deliberation; Costa & McCrae, 1995). Dudley, Orvis, Lebiecki, and Cortina (2006) showed that the facets of conscientiousness explained additional variance of 1 to 24% on job performance beyond the global trait conscientiousness. A meta-analysis by Schmidt and Hunter (1998) revealed incremental validity of conscientiousness beyond intelligence of about 10% additional explained variance. Thus far, however, there are no studies analyzing concurrent validity of currently used intelligence tests for career success in German-speaking countries, especially by looking at fluid and crystallized intelligence separately and additionally analyzing facets of conscientiousness. This dissertation aimed to fill these gaps: Concurrent validity of the RIAS for career success was analyzed, differentiating general intelligence as fluid and crystallized intelligence. Moreover, the incremental validity of conscientiousness and its narrow facets beyond intelligence was examined.



### 3. Research Questions

The present dissertation had two aims: (1) establishing the construct validity, in particular factorial validity, of the RIAS, a currently used intelligence test for individuals aged 3 to 99 years; and (2) analyzing the criterion validity of currently used intelligence tests in German-speaking countries for scholastic achievement and career success. Figure 1 offers a schematic overview of the dissertation concept. In particular, this dissertation addressed the following research questions:

1. Construct validation of the Reynolds Intellectual Assessment Scales (RIAS):
  - a. Are the single-factor structure and the two-factor structure of the RIAS identical across different language groups? (Study 1)
  - b. Are the single-factor structure and the two-factor structure of the RIAS identical across individuals with and without a migration background? (Study 2)
  - c. Do individuals with and without a migration background show mean differences in the latent factors of the RIAS? (Study 2)
2. Criterion validation of currently used intelligence tests in German-speaking countries:
  - a. Is intelligence, measured by the IDS, RIAS, SON-R 6–40, and WISC-IV, a valid predictor of longitudinal scholastic achievement in typically developing children? (Study 3)
  - b. Is intelligence, in particular crystallized intelligence, measured by the RIAS, a valid predictor of concurrent career success? (Study 4)
  - c. Can conscientiousness and its facets explain additional variance in career success beyond intelligence? (Study 4)



*Figure 1.* Dissertation concept. RIAS = Reynolds Intellectual Assessment Scales; numbers in parentheses refer to the studies that make up this dissertation: Study 1 (Gygi, Ledermann, Grob, & Hagmann-von Arx, *submitted*), Study 2 (Gygi, Fux, Grob, & Hagmann-von Arx, 2016), Study 3 (Gygi, Hagmann-von Arx, Schweizer, & Grob, *submitted*), Study 4 (Hagmann-von Arx, Gygi, Weidmann, & Grob, 2016).

## 4. Method

The following chapter gives an overview on the studies and samples that make up this dissertation, followed by outlines of the measures and statistical analyses of the studies.

### 4.1 Studies and Samples

**Study 1 (Gygi, Ledermann, Grob, & Hagmann-von Arx, *submitted*).** The objective of Study 1 was to analyze the single-factor and two-factor structure of the RIAS as well as its measurement invariance across all four RIAS language versions. Therefore, the standardization samples of the English ( $n = 2,438$ ), Danish ( $n = 983$ ), German ( $n = 2,103$ ), and Spanish ( $n = 1,933$ ) RIAS version were used. Altogether, they included a total of 7,457 individuals aged 3 to above 90 years. The samples were evenly distributed with respect to sex and were intended to match the census data on educational attainment of the respective countries. Subjects in the English sample were from the United States, subjects in the Danish sample were from Denmark, subjects in the German sample were from Switzerland and from Germany, and subjects in the Spanish sample were from Spain.

**Study 2 (Gygi, Fux, Grob, & Hagmann-von Arx, 2016).** The aim of Study 2 was to assess measurement invariance and latent mean differences in the German version of the RIAS across individuals with and without a migration background. The sample was a subsample of the German RIAS standardization and included 632 individuals with ( $n = 316$ ) and without ( $n = 316$ ) a migration background. Individuals were 3 to 99 years old ( $M_{\text{age}} = 15.79$  years,  $SD = 16.81$ ; 48% females, 52% males). For individuals without a migration background, German was their first language, whereas for individuals with a migration background German was their second language. First languages of individuals with a migration background were, for example, Turkish (21%), Italian (8%), Serbian (6%), and Spanish (5%).

**Study 3 (Gygi, Hagmann-von Arx, Schweizer, & Grob, *submitted*).** The aim of Study 3 was to analyze the predictive power of four currently used intelligence tests (i.e., IDS, RIAS, SON-R 6–40, WISC-IV) in German-speaking countries for longitudinal scholastic achievement (i.e., school grades) in typically developing children. Therefore, a subsample of the German RIAS standardization sample was analyzed. The sample included 103 children aged 6 to 11 years at Study Wave 1 ( $M_{\text{age}} = 9.18$  years,  $SD = 0.93$ , 52% females, 48% males) and 54 children aged 10 to 13 years at Study Wave 2, three years later ( $M_{\text{age}} = 11.77$  years,  $SD = 0.79$ , 52% females, 48% males).

**Study 4 (Hagmann-von Arx, Gygi, Weidmann, & Grob, 2016).** Study 4 had two objectives: To analyze the concurrent validity of intelligence (fluid, crystallized) for extrinsic (i.e., occupational skill level, income) and intrinsic (i.e., job satisfaction) career success; and to analyze the incremental validity of conscientiousness and its facets predicting career success beyond intelligence. The sample included 121 adults from the German RIAS standardization aged 21 to 77 years ( $M_{\text{age}} = 48.45$  years,  $SD = 12.54$ ; 60% females, 40% males).

### 4.3 Measures

**Studies 1 and 2.** For Studies 1 and 2, intelligence was assessed through the RIAS. The RIAS is an individually administered intelligence test that assesses a Composite Intelligence Index (CIX), which can be further divided into a Verbal Intelligence Index (VIX) and a Nonverbal Intelligence Index (NIX). The VIX and NIX consist of two subtests each. The verbal subtests measure vocabulary and verbal reasoning and the nonverbal subtests measure abstract reasoning. The CIX ( $M = 100$ ,  $SD = 15$ ) is calculated from the sum of the  $T$  scores ( $M = 50$ ,  $SD = 10$ ) of the four intelligence subtests. The RIAS also includes a conormed, supplemental Composite Memory Index (CMX) consisting of two subtests that measure verbal and nonverbal memory, respectively. The memory subtests were not included in the

analyses of this dissertation. Reliability for the German version is high with Cronbach's  $\alpha = .94$  for the VIX,  $.95$  for the NIX, and  $.96$  for the CIX. Reliabilities for the other three RIAS language versions are also high with Cronbach's alphas ranging from  $.93$  to  $.99$  for the VIX, from  $.90$  to  $.98$  for the NIX, and from  $.95$  to  $.99$  for the CIX.

**Study 3.** For Study 3, intelligence was assessed with the IDS, the RIAS, the SON-R 6–40, and the WISC-IV at Study Wave 1. Scholastic achievement (i.e., school grades) was assessed based on parental reports at Study Wave 2.

The IDS (Grob et al., 2013) assesses general intelligence and was standardized in 2007 and 2008 in Switzerland, Germany, and Austria for children aged 5 to 10 years. The IDS intelligence index ( $M = 100$ ,  $SD = 15$ ) consists of seven subtests that measure visual perception, selective attention, reasoning (abstract and figural), working memory (phonological and visual-spatial), and long-term memory and thus assesses primarily fluid intelligence. Additional to intelligence, another five developmental domains can be assessed: Psychomotor skills, social–emotional competence, language, mathematics, and achievement motivation. For this dissertation, only the intelligence index was analyzed. Reliability for the IDS intelligence index is high with Cronbach's  $\alpha = .92$ .

The SON-R 6–40 (Tellegen et al., 2012) is an individually administered intelligence test that assesses intelligence for individuals aged 6 to 40 years nonverbally. It was standardized in Germany and the Netherlands and comprises four subtests that measure abstract reasoning and visual–spatial processing, thus representing fluid intelligence. Reliability for the SON-R 6–40 composite intelligence index ( $M = 100$ ,  $SD = 15$ ) is high with Cronbach's  $\alpha = .95$ .

The WISC-IV (Wechsler, 2003) is an individually administered intelligence test measuring general intelligence for children and adolescents aged 6 to 16 years. Originally developed and standardized in the United States, it has been translated, adapted, and standardized worldwide as, for example, for Austria, Germany, and Switzerland (Petermann

& Petermann, 2011). The WISC-IV includes 10 core subtests that measure general intelligence, verbal comprehension, working memory, processing, and reasoning. Thus, the WISC-IV provide the Full Scale IQ (FSIQ), and four specific intelligence indices ( $M = 100$ ;  $SD = 15$ ): The Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index. Reliability for the WISC-IV is high, with  $r = .97$  for the FSIQ and  $r = .87$  to  $.94$  for the specific intelligence indices.

Children's scholastic achievement was assessed in mathematics and language (German). Parents reported their child's standardized school grades on a scale from 1 (*poorest grade*) to 6 (*best grade*; grades 4–6 represent the passing range) based on the school records of the latest term (overall grades).

**Study 4.** In Study 4, intelligence was assessed through the RIAS. Conscientiousness was assessed through the personality questionnaire NEO-Personality Inventory-Revised (NEO-PI-R; Ostendorf & Angleitner, 2004). Career success was assessed through questions about the participants' profession, income, and job satisfaction.

The German version of the NEO-PI-R (Ostendorf & Angleitner, 2004) is a self-report inventory consisting of 240 items that measure the five major dimensions of personality with 30 items each (i.e., neuroticism, extraversion, openness, agreeableness, conscientiousness). Each domain includes six facets, which are assessed through 8 items each. Responses are made on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*). In this dissertation, the NEO-PI-R was applied to assess conscientiousness and its facets (competence, order, dutifulness, achievement striving, self-discipline, deliberation). Reliability for the composite score of conscientiousness is high with  $\alpha = .93$  and moderate to high for the facets of conscientiousness with  $\alpha = .71$  to  $.84$ .

Career success was assessed objectively (occupational skill level and income) and subjectively (job satisfaction): First, participants were asked about their profession, whereupon their occupational skill level was determined. This information was then encoded

according to four skill levels from 1 (*unskilled*) to 4 (*highly skilled*) as defined in the International Standard Classification of Occupations (ISCO-08; International Labour Organization, 2008). Second, participants were given one item (“How high is your gross income per year?”) to assess their income. Participants’ income was divided by their country’s most recent purchasing power parity to control for cross-country differences in income, as participants from Switzerland and Germany were included. Income was log-transformed for the analyses. Finally, job satisfaction was assessed with a German self-report survey (Neuberger & Allerbeck, 1978) containing eight items. The items were rated on a 5-point scale from 1 (*does not apply at all*) to 5 (*applies completely*). In this dissertation, Cronbach’s alpha was moderate with  $\alpha = .76$ .

#### 4.4 Statistical Analyses

**Studies 1 and 2.** In Studies 1 and 2, single-group and multi-group confirmatory factor analyses (CFAs) were conducted using SPSS Amos 21 to assess the RIAS factor structure and its measurement invariance across different language groups (Study 1) and individuals with and without a migration background (Study 2). Measurement invariance was conducted according to Meade, Johnson, and Braddy (2008), Meredith (1993), and Widaman and Reise (1997).

**Studies 3 and 4.** For Studies 3 and 4, multiple regression analyses were carried out using SPSS 22 to assess the predictive validity of intelligence for scholastic achievement (Study 3) and career success (Study 4). Additionally in Study 4, the incremental validity of conscientiousness and its facets for career success were assessed through multiple regression analysis. Finally, for Study 4, bootstrap procedures were applied for all analyses, as the distributions of some variables showed deviations from normality (Chernick, 2008).

## 5. Synopsis of Results

The following chapter summarizes the results of the studies included in this dissertation for the research questions laid out in Chapter 3.

### 5.1 Factorial Validity of the RIAS

Study 1 (Gygi, Ledermann et al., *submitted*) analyzed the RIAS factor structures for all four language versions of the RIAS and additionally examined measurement invariance across groups. Results of Study 1 revealed that both the single-factor and the two-factor structure of the RIAS were identified for each of the four language groups (i.e., English, Danish, German, Spanish). Fit indices suggested that the two structures were empirically equivalent in the single-group CFAs. Further, a multi-group CFA revealed scalar measurement invariance across groups for both structures, indicating identical factor structures for the four RIAS versions. When the three recently standardized RIAS versions (i.e., Danish, German, Spanish) were separately compared with the original English RIAS version, results showed full measurement invariance for the English and German RIAS versions. In contrast, the Danish and Spanish RIAS versions differed significantly from the English RIAS version in terms of residual measurement invariance.

Study 2 (Gygi, Fux et al., 2016) analyzed the factor structure of the RIAS and its measurement invariance across individuals with and without a migration background. Results revealed that the single-factor and the two-factor structure of the RIAS were supported for both groups. Fit indices revealed the two structures to be empirically equivalent. Further, scalar measurement invariance was found for the two-factor structure across groups. The single-factor structure showed partial scalar measurement invariance, indicating the subtest OIO to be the major contributor of noninvariance. Latent variances and latent means could be analyzed only for the two-factor structure, as scalar invariance is a prerequisite for analyzing latent variances and means (Meredith, 1993). Results showed that latent variances of verbal intelligence were invariant across groups, whereas latent variances of nonverbal intelligence



differed across groups (individuals with a migration background showed a wider range of scores). Further, latent means of verbal and nonverbal intelligence differed between groups and were lower for individuals with a migration background compared with individuals without a migration background. Moreover, effect sizes indicated a large effect for verbal intelligence and a small effect for nonverbal intelligence.

## **5.2 Concurrent and Predictive Validity of Currently Used Intelligence Tests**

Study 3 (Gygi, Hagmann-von Arx et al., *submitted*) analyzed the predictive validity of currently used intelligence tests for school grades in mathematics and language longitudinally. Results revealed that the composite intelligence indices of four currently used intelligence tests (IDS, RIAS, SON-R 6–40, WISC-IV) in German-speaking countries predicted averaged school grades in typically developing school-aged children. Moreover, IDS and SON-R 6–40 were significant positive predictors for school grades in mathematics, while IDS and RIAS were significant positive predictors for school grades in language. The WISC-IV did not show significant associations with mathematics and language.

Study 4 (Hagmann-von Arx, Gygi et al., 2016) examined the concurrent validity of fluid and crystallized intelligence for career success and additionally analyzed incremental validity of conscientiousness for career success. Study 4 showed that crystallized, but not fluid intelligence (i.e., the RIAS VIX but not the NIX) predicted concurrent career success: RIAS VIX predicted occupational skill level but not income and job satisfaction. However, the association of intelligence and job satisfaction was moderated by occupational skill level. Thus, the RIAS VIX negatively predicted job satisfaction at lower occupational skill levels, but there was no association in higher occupational skill levels. Further, the results of Study 4 revealed that conscientiousness explained variance in income and job satisfaction but not in occupational skill level. Regarding the facets of conscientiousness, results showed that the more self-disciplined people reported higher incomes, whereas people with higher competence and achievement striving were more satisfied with their jobs.

## 6. General Discussion

The aim of the present dissertation was to expand current knowledge on intelligence assessment and its construct and criterion validity by examining the factor structure of the RIAS and its measurement invariance across different groups of individuals, as well as by analyzing the concurrent and predictive validity of intelligence tests for scholastic achievement and career success. Additionally, the incremental validity of conscientiousness for career success beyond intelligence was examined.

### 6.1 Factorial Validity of the RIAS

In Studies 1 and 2 (Gygi, Ledermann et al., *submitted*; Gygi, Fux et al., 2016), the construct validity of the RIAS was analyzed. CFAs were conducted to examine its factor structure for different groups of individuals (i.e., individuals from different language groups and individuals with and without a migration background, respectively) and to assess measurement invariance across these groups. Results supported a two-factor and a single-factor structure for each RIAS language version as well as for individuals with and without a migration background. These findings are in line with previous studies that found a two-factor structure (Beaujean et al., 2009; Hagmann-von Arx & Grob, 2014; Hartmann & Andresen, 2011; Reynolds & Kamphaus, 2003; Santamaría & Fernández Pinto, 2009) as well as with studies that found a single-factor structure (Beaujean & McGlaughlin, 2014; Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007).

Separate measurement invariance analyses of the English and German RIAS versions revealed full measurement invariance, whereas the Danish and Spanish versions differed significantly from the English version in their residual variance. This indicates that English and German RIAS versions have comparable relationships of observed scores and their latent constructs (Meredith, 1993; Milfont & Fischer, 2010). Differences in residual variance suggest variations in reliabilities of the Danish and Spanish versions compared with the

English version (DeShon, 2004). Across the four RIAS language versions, scalar measurement invariance was established. Hence, the relationship of the observed scores and their latent constructs is comparable across the four language groups, while possible variations in reliabilities led again to differences in residual variances.

Across individuals with and without a migration background, only the two-factor structure showed scalar measurement invariance, suggesting that the relationship between observed scores and latent constructs was the same across groups (Meredith, 1993; Milfont & Fischer, 2010). Potential variation in the scales' reliabilities across groups led to the noninvariance in the residuals (DeShon, 2004). Furthermore, noninvariance on the scalar invariance level in the single-factor structure indicates a different relationship between the observed scores and the latent construct across individuals with and without a migration background. The lack of scalar invariance permits a comparison of latent variances and means (Meredith, 1993). However, partial scalar invariance was found, with the subtest OIO being a major contributor to differences across groups. This is in line with findings of previous studies in which the subtest OIO led to decreased model fit in the English RIAS version (Beaujean et al., 2009; Beaujean & McGlaughlin, 2014). Hence, group differences in the latent factor *g* can be explained either through mean differences in the subtest OIO or through true mean differences in the latent factor (Meredith & Teresi, 2006). Taken together, these results suggest that practitioners should focus primarily on the VIX and NIX and not on the CIX when comparing RIAS test results of individuals with and without a migration background.

Results for individuals with a migration background revealed lower latent means compared with individuals without a migration background, especially on the VIX but also on the NIX. This is in line with studies that also found lower means in intelligence test scores for individuals with a migration background (Calero, Fernández-Parra et al., 2013; Hagmann-von Arx et al., 2013; Harris et al., 2008; Resing et al., 2009). Moreover, larger effect sizes in the

differences on the VIX may indicate language differences in individuals with a migration background. However, environmental and cultural distinctions may have also been influential (Calero, Mata et al., 2013; Daseking et al., 2008; Georgas et al., 2003; Hagmann-von Arx et al., 2013; Harris et al., 2008; Mackintosh, 2011; Resing et al., 2009; Schölmerich & Leyendecker, 2009). Studies found, for example, that familiarity with test tasks (Calero, Mata et al., 2013; Resing et al., 2009), stereotype threat (Appel, Weber, & Kronberger, 2015; Baltes & Rudolph, 2010), and parents' language proficiency (Lahaie, 2008) can lead to lower test performance, and this may also have contributed to lower test scores in the VIX and the NIX in the present dissertation. Hence, results of the current study indicate that examiners must consider migration background of an examinee when assessing intelligence and when comparing test results of individuals with and without a migration background. One approach, for example, is the "dynamic" testing method, which may be used instead of the typically used "static" testing method (Resing et al., 2009, p. 445). In the dynamic testing method, tasks are first introduced to familiarize an examinee with their nature. In a subsequent round, task improvements are assessed to estimate learning potential of the examinee. Studies showed that dynamic testing leads to decreasing test score differences between individuals with and without a migration background (Calero, Mata et al., 2013; Resing et al., 2009). Practitioners should therefore choose dynamic over static testing methods when testing individuals with a migration background, so that effects of familiarity with intelligence test tasks are minimized. Also, future studies should examine whether benefits of the dynamic testing approach are also present when assessing individuals with a migration background using the German version of the RIAS.

## **6.2 Concurrent and Predictive Validity of Currently Used Intelligence Tests**

Studies 3 and 4 (Gygi, Hagmann-von Arx et al., *submitted*; Hagmann-von Arx, Gygi et al., 2016) assessed the concurrent and predictive validity of intelligence for scholastic

achievement and career success. Moreover, Study 4 analyzed incremental validity of conscientiousness and its facets for career success.

Regarding scholastic achievement, the assessment of intelligence using the IDS, RIAS, SON-R 6–40, and WISC-IV significantly predicted averaged school grades across 3 years with moderate effect sizes. These results are in accordance with previous studies showing that intelligence is a positive and moderate to strong predictor of academic achievement (e.g., Deary et al., 2007; Gut et al., 2013; Roth et al., 2015). When analyzing subject domains separately (i.e., mathematics, language), results revealed that the IDS and SON-R 6–40 general intelligence predicted school grades in mathematics with small to moderate effect sizes. The IDS assesses working memory in four out of seven subtests (i.e., selective attention, phonological memory, visual-spatial memory, auditory memory), while the SON-R 6–40 primarily assesses visual-spatial abilities (i.e., analogies, mosaic, draw pattern). Hence, results are in line with previous findings showing that particularly working memory (Dehn, 2008; Raghubar, Barnes, & Hecht, 2010) and visual-spatial abilities (Verdine, Golinkoff, Hirsh-Pasek, Newcombe, Filipowicz, & Chang, 2014; Wai, Lubinski, & Benbow, 2009) are associated with mathematics. The RIAS and the WISC-IV did not significantly predict school grades in mathematics, though the measured small effect sizes were positive. This contradicts previous studies that found general intelligence to be a moderate to strong predictor of school grades in mathematics (Roth et al., 2015). However, the RIAS does not include subtests that measure working memory and visual-spatial abilities, while the WISC-IV includes perceptual reasoning and working memory in one specific index each out of four specific indices. Thus, the RIAS and WISC-IV measure working memory and visual-spatial abilities to a lesser extent, which may have weakened the relationship with mathematics. In this vein, it is plausible that the IDS (that primarily measures working memory) and SON-R 6–40 (that primarily measures visual-spatial abilities) in particular, were significantly associated with school

grades in mathematics, as working memory and visual-spatial abilities have been found to be associated with mathematics (e.g., Dehn, 2008; Verdine et al., 2014).

Regarding the prediction of language, the IDS and the RIAS general intelligence moderately predicted school grades. The relationship between IDS and school grades in language is in line with previous literature supporting a moderate to strong association of working memory and language (e.g., Dehn, 2008). The association of the RIAS and school grades in language is a coherent finding as two out of four subtests measure verbal reasoning and verbal abilities. The SON-R 6–40 and the WISC-IV showed no significant associations with language, although small effect sizes were positive. This stands in contrast to previous studies that found a moderate association of intelligence and school grades in language (Roth et al., 2015). However, in contrast to the IDS, the SON-R 6–40 does not measure working memory, which has been shown to be moderately associated with language (e.g., Dehn, 2008). Furthermore, the SON-R 6–40 focuses only on nonverbal intelligence and does not include subtests measuring verbal abilities as with the RIAS. The WISC-IV, however, includes a specific index each for verbal comprehension and working memory. It might be possible that in our study, these two out of four specific indices were not sufficient to significantly explain variance in school grades in language.

Notably, effect sizes were only small to moderate, and thus, somewhat lower than expected, according to meta-analytical results by Roth et al. (2015). However, effects may be smaller in a single study, as seen, for example, in Gut et al. (2012, 2013). Furthermore, the sample used showed somewhat higher general intelligence scores than the general population and had a narrow range in school grades, which were all in the passing range. This may have weakened correlations between intelligence and school grades here, due to range restrictions (Roth et al., 2015; Sternberg et al., 2001). Moreover, the predictive validity of intelligence tests on school grades was analyzed for primary school children. Roth et al. (2015) found smaller correlations between intelligence and school grades in primary school-level students

compared to higher grade-level students. These authors claim that lower intelligence scores in primary school might be more easily compensated for with practice than in higher grade levels, which may also have led to smaller effect sizes in Study 3 (Gygi, Hagmann-von Arx et al., *submitted*).

However, a large portion of variance in scholastic achievement remains that cannot be explained through intelligence. Rost (2009) and Wentzel (1991) state that school grades are influenced not only by cognition but also by social competence or self-control, as learning takes place in a social context and requires persistence over an extended period of time. This points to other constructs that additionally predict scholastic achievement, such as school engagement (Reyes, Brackett, Rivers, White, & Salovey, 2012), motivation (Steinmayr & Spinath, 2009), self-control (Duckworth, Quinn, & Tsukayama, 2012), personality (Poropat, 2009; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007), and social-emotional competencies (Gut et al., 2012; Oberle, Schonert-Reichl, Hertzman, & Zumbo, 2014), as well as theory of intelligence (Blackwell, Trzesniewski, & Dweck, 2007).

Therefore, Study 4 (Hagmann-von Arx, Gygi et al., 2016) additionally analyzed the effect of conscientiousness beyond intelligence, in predicting career success. Results revealed that intelligence assessed by the RIAS, and in particular crystallized intelligence, was a significant predictor of occupational skill level, which is in line with previous findings that intelligence increases with increasing occupational skill level (e.g., Schmidt & Hunter, 2004) as well as with Schmidt's (2014) theoretical model. However, effect sizes for the association of intelligence and occupational skill level were smaller than expected. A reason for this may be because of range restrictions (Sternberg et al., 2001), as the variances in IQ scores were lower than expected ( $SD < 15$ ). Further, intelligence showed no association with income, which is in contrast to previous studies (Ng et al., 2005; Strenze, 2007). However, the present sample size might have been too small to detect the expected small effect sizes, as earlier studies also found only small correlations ( $r = .16$  to  $.23$ ) between intelligence and income

(Ng et al., 2005; Strenze, 2007). Furthermore, results indicated a moderation of the association between crystallized intelligence and job satisfaction by occupational skill level: Crystallized intelligence negatively predicted job satisfaction in lower occupational skill levels, whereas no association was found in higher occupational skill levels. This is in line with previous research (Ganzach & Fried, 2012; Lounsbury et al., 2004) and indicates the importance of a good person–environment fit for high job satisfaction. Also, practitioners may place more importance on crystallized intelligence than fluid intelligence when making predictions for career success.

Moreover, results of Study 4 (Hagmann-von Arx, Gygi et al., 2016) revealed additional explained variance of 12 to 13% in income and job satisfaction by conscientiousness. These results are in line with previous findings supporting the incremental validity of conscientiousness on career success (Judge, Heller, & Mount, 2002; Ng et al., 2005). Also, it conforms to Schmidt's theoretical model (2014), which states that conscientiousness explains additional variance of career success. This indicates that people who are more conscientious tend to report higher income and job satisfaction independent of their intelligence scores. However, conscientiousness did not explain additional variance in occupational skill level, which stands in contrast to findings of Hurtz and Donovan (2000). Finally, the analysis of specific facets of conscientiousness and their association with career success revealed that more self-disciplined individuals reported higher income, whereas more competent and achievement-striving individuals reported higher job satisfaction. Thus, our results are consistent with previous research (Dudley et al., 2006; Judge, Rodell, Klinger, Simon, & Crawford, 2013) and emphasize the importance of assessing not only conscientiousness but also its narrow facets when making predictions regarding career success.

In sum, our results revealed predictive validity of the IDS, RIAS, SON-R 6–40, and WISC-IV in German-speaking countries for scholastic achievement (Neukrug & Fawcett,



2015). Furthermore, our results support the concurrent validity of the RIAS and incremental validity of conscientiousness for career success.

### **6.3 Strengths and Limitations**

This dissertation has several strengths. First, the analysis of the RIAS across four language groups is a clear strength as thereby valid interpretations across these four RIAS language versions can be supported. Accordingly, our results support the use of, for example, the Spanish RIAS for people emigrating from Spain to Switzerland when language proficiency is lacking. This is important as mobility increases and a growing number of individuals are seeking psychological assessment in a foreign language environment.

Second, the thorough examination of the German RIAS with different subsamples contributes to the validation of its scores and their interpretations. Intelligence tests must be validated through various studies (Neukrug & Fawcett, 2015). Also, intelligence tests are seen to increase in validity and reliability with increasing length. Yet, while the RIAS is highly time-efficient, its validity could be supported in this dissertation through several studies: Different aspects of validity were analyzed for the RIAS, namely, construct validity, which includes the analyses of factorial validity, and criterion validity, which includes the assessment of concurrent and longitudinal predictive power of the RIAS. Thus, the use of the RIAS in research as well as in practice, is again supported.

Further, this dissertation supported the validity of the RIAS across individuals with and without a migration background. This is of increasing interest as migration is rising worldwide (Eurostat, 2014; U.S. Department of Homeland Security, 2014) and more individuals with a migration background undergo psychological assessment. The results described herein should give practitioners confidence in the use of currently available intelligence tests, and signal a need for considering language barriers and environmental and cultural differences due to migration background in the interpretation of test results.

This study also examines the predictive power of several currently used intelligence tests for longitudinal scholastic achievement and concurrent career success, in German-speaking countries. Hence, practitioners should feel confident in the application of the respective intelligence tests for questions that typically arise in school and counseling psychology. Moreover, the demonstrated validity of these tests lend support to the interpretations drawn from their results.

Finally, not only the predictive power of intelligence, but also the incremental validity of conscientiousness was considered. Hence, this dissertation analyzed factors beyond intelligence, which are predictors of achievement. Additionally, conscientiousness was analyzed both as a composite trait and dissected according to its narrow facets. In this way, divergent relationships were detected between the six facets of conscientiousness and achievement.

In parallel to its numerous strengths, this dissertation also has some limitations. First, while a clear strength of administering the RIAS is its brevity, this also limits the possibility of statistically evaluating alternative and concurrent factor structures. Similarly, the performance of fit indices in CFAs with small degrees of freedom is not known (e.g., Kenny, Kaniskan, & McCoach, 2015). However, Kenny et al. (2015) has stated that the use of large sample sizes mitigates the problem of small degrees of freedom, at least for the root mean square error of approximation.

Further, only standardized scores and not raw scores of the English, Danish, and Spanish RIAS standardization samples were available, as not all test publishers made their data available. Hence, we could not analyze the factor structure and its measurement invariance using raw scores or on item level. Future studies could analyze the factor structure and its measurement invariance using raw scores. This would enable researchers to examine additional questions, such as whether general populations differ in their factor means or if items correlate with the subtest in a similar manner across groups. However, publishers need

to make raw test data accessible in the spirit of open science (Albagli, Maciel, & Abdo, 2015). Also, as the RIAS covers the full life span, raw scores are strongly correlated with age (Reynolds & Kamphaus, 2003). Consequently, analyses using raw scores would necessitate the assessment of age-homogeneous groups.

Further, our sample of individuals with a migration background in Study 2 (Gygi, Fux et al., 2016) was determined by language spoken (native speaker vs. speaker of a foreign language). Yet migration background can be operationalized in different ways. For example, future studies might additionally take into account how long individuals and their families have been living in the country of immigration. Furthermore, the languages of individuals with a migration background varied greatly. Although this diversity characterizes Switzerland and Germany (Bundesamt für Statistik, 2015; Statistisches Bundesamt, 2013), these individuals do not represent a homogeneous group. Future research could thus focus on specific groups, such as individuals from collectivist or individualist cultures (e.g., Hofstede, 2001) or individuals with first languages in the same language family (e.g., Ruhlen, 1991).

Additionally, while sample sizes of Study 1 (Gygi, Ledermann et al., *submitted*) and Study 2 (Gygi, Fux et al., 2016) were large and can be considered a strength of these studies, Study 3 (Gygi, Hagmann-von Arx et al., *submitted*) and Study 4 (Hagmann-von Arx, Gygi et al., 2016), had rather small sample sizes. Consequently, in Studies 3 and 4, it was not possible to detect small effects or to generalize results to the population or to individuals with special needs.

Finally, the studies in this dissertation used subsamples of standardization samples and therefore analyzed only typically developed individuals. Hence, results of this dissertation cannot be generalized to clinical samples or to individuals with special needs. Accordingly, future studies might replicate present findings using clinical samples similar to others (e.g., Beaujean & McGlaughlin, 2014; Canivez et al., 2014; Mayes & Calhoun, 2007).

## **6.4 Conclusion and Outlook**

In conclusion, results of this dissertation support the factor structure of the RIAS and its measurement invariance across different groups of individuals. As Messick (1995) noted, the appropriateness of test scores and their interpretation must be supported by statistical evidence. The construct validity of the RIAS can be supported and therefore the interpretation of RIAS test scores may be considered reliable. Further, this dissertation revealed latent mean differences suggesting a disadvantage of individuals with a migration background, in particular in verbal intelligence. Consequently, practitioners must consider the migration background of an examinee, as language and cultural differences may lead to lower test scores in individuals with a migration background compared to those without a migration background, and subsequently may result in missed educational opportunities (Calero, Fernández-Parra et al., 2013). As migration is rising (Eurostat, 2014; U.S. Department of Homeland Security, 2014), an increasing number of individuals with a migration background will likely undergo psychological assessment. Therefore, future research should further investigate the influence of language, environment, and culture on intelligence test results, as individuals with a migration background in this dissertation scored lower not only in verbal intelligence, but also in nonverbal intelligence.

This dissertation also supports the predictive power of the IDS, RIAS, SON-R 6–40, and WISC-IV for longitudinal scholastic achievement, as well as for concurrent career success. These results should give practitioners confidence in the application of currently used intelligence tests when making predictions regarding scholastic achievement and career success. However, a large portion of achievement variance cannot be explained by intelligence, highlighting the role of additional influential factors. In the present dissertation, the importance of including conscientiousness in the association with intelligence and achievement has been shown, by revealing the incremental validity of conscientiousness in career success. Future studies could also focus on other individual characteristics, such as

self-control (Duckworth et al., 2012) and social-emotional competencies (Gut et al., 2012; Oberle, Schonert-Reichl, Hertzman, & Zumbo, 2014). Moreover, future studies might also consider ecological systems theory (Bronfenbrenner, 1979): In this vein, not only individual characteristics but also the influence of contextual factors on scholastic achievement would be analyzed, such as, for example, the influence of parents, peers, school environment, social policies, or society (e.g., Bandura, Barbaranelli, Caprara, & Pastorelli, 2001; Burke & Sass, 2013; Flere, Krajnc, Klanjšek, Musil, & Kirbiš, 2010). Finally, future studies might analyze the predictive power of intelligence and other relevant factors for achievement longitudinally across the entire life span. Previous studies, for instance, have analyzed the association between intelligence and several health outcomes over the period of 20 to 65 years, using the Scottish Mental Survey of 1932 and the U.S. National Longitudinal Study of Youth, respectively (Deary, Whiteman, Starr, Whalley, & Fox, 2004; Der, Batty, & Deary, 2014). Hence, one could analyze not only predictive power over the long haul, but also stability of the predictors, which in turn may reveal how changes in predictors influence achievement as well as the direction of the effects.

In sum, the present dissertation extends current knowledge of intelligence assessment and supports construct and criterion validity of currently used intelligence tests in German-speaking countries. Nevertheless, factors beyond intelligence must also be considered when making predictions regarding scholastic achievement and career success.

### References

- Albagli, S. Maciel, M. L., & Abdo A. H. (2015). *Open science, open issues*. Brasilia, Brasil: IBICT; Rio de Janeiro, Brasil: Unirio.
- Andrews, J. J. W. (2007). RIAS: Reynolds Intellectual Assessment Scales. *Journal of Psychoeducational Assessment*, 25, 402-408. doi:10.1177/0734282907300381
- Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology*, 6, 900. doi:10.3389/fpsyg.2015.00900
- Baltes, B. B., & Rudolph, C. W. (2010). Examining the effect of negative Turkish stereotypes on evaluative workplace outcomes in Germany. *Journal of Managerial Psychology*, 25, 148-158. doi:10.1108/02683941011019357
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72, 187-206. doi:10.1111/1467-8624.00273
- Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, 17, 278-288. doi:10.1016/j.annepidem.2006.07.010
- Beaujean, A. A., & McGlaughlin, S. M. (2014). Invariance in the Reynolds Intellectual Assessment Scales for Black and White referred students. *Psychological Assessment*, 26, 1394. doi:10.1037/pas0000029
- Beaujean, A. A., McGlaughlin, S. M., & Margulies, A. S. (2009). Factorial validity of the Reynolds Intellectual Assessment Scales for referred students. *Psychology in the Schools*, 46, 932-950. doi:10.1002/pits.20435
- Binet, A., & Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Anée Psychologique*, 11, 191-336.

- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246-263. doi:10.1111/j.1467-8624.2007.00995.x
- Borsuk, E. R., Watkins, M. W., & Canivez, G. L. (2006). Long-term stability of membership in a Wechsler Intelligence Scale for Children—Third edition (WISC-III) subtest core profile taxonomy. *Journal of Psychoeducational Assessment, 24*, 52-68. doi:10.1177/0734282905285225
- Braden, J. P., & Niebling, B. C. (2005). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 615-630). New York, NY: Guilford Press.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bundesamt für Statistik. (2015). *Migration und Integration* [Migration and integration]. Retrieved from <http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/07/blank/key/01/01.htmlx>
- Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics, 31*, 51-82. doi:10.1086/666653
- Calero, M. D, Fernández-Parra, A., López-Rubio, S., Carles, R., Mata, S., Vives, C., Navarro, E., et al. (2013). Variables involved in personal, social and school adjustment in a sample of preschool-aged children from different cultural backgrounds. *European Journal of Psychological Education, 28*, 133-155. doi:10.1007/s10212-012-0107-8
- Calero, M. D., Mata, S., Carles, R., Vives, C., López-Rubio, S., Fernández-Parra, A., & Navarro, E. (2013). Learning potential assessment and adaptation to the educational context: The usefulness of the ACFS for assessing immigrant preschool children. *Psychology in the Schools, 50*, 705-721. doi:10.1002/pits.21701

- Canivez, G. L., Watkins, M. W., James, T., Good, R., & James, K. (2014). Incremental validity of WISC-IV<sup>UK</sup> factor index scores with a referred Irish sample: Predicting performance on the WIAT-II<sup>UK</sup>. *British Journal of Educational Psychology*, 84, 667-684. doi:10.1111/bjep.12056
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Amsterdam, the Netherlands: North-Holland.
- Chernick, M. R. (2008). *Bootstrap methods: A guide for practitioners and researchers*. Hoboken, NJ: Wiley.
- Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment*, 64, 21-50. doi:10.1207/s15327752jpa 6401\_2
- Daseking, M., Lipsius, M., Petermann, D., & Waldmann, H.-C. (2008). Differenzen im Intelligenzprofil bei Kindern mit Migrationshintergrund: Befunde zum HAWIK-IV. [Application of the German WISC-IV in children with diverse cultural and linguistic backgrounds.] *Kindheit und Entwicklung*, 17, 76-89. doi:10.1026/0942-5403.17.2.76
- Deary, I. J. (2009). Introduction to the special issue on cognitive epidemiology. *Intelligence*, 37, 517-519. doi:10.1016/j.intell.2009.05.001
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, 63, 453-482. doi:10.1146/annurev-psych-120710-100353
- Deary, I. J. (2014). The stability of intelligence from childhood to old age. *Current Directions in Psychological Science*, 23, 239-245. doi:10.1177/0963721414536905
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21. doi:10.1016/j.intell.2006.02.001
- Deary, I. J., Pattie, A., & Starr, J. M. (2013). The stability of intelligence from age 11 to age



- 90 years The Lothian Birth Cohort of 1921. *Psychological Science*, 24, 2361–2368.  
doi:10.1177/0956797613486487
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, 86, 130-147. doi:10.1037/0022-3514.86.1.130
- Dehn, M. J. (2008). *Working memory and academic learning. Assessment and intervention*. Hoboken, NJ: Wiley.
- Der, G., Batty, G. D., & Deary, I. J. (2009). The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence*, 37, 573-580. doi:10.1016/j.intell.2008.12.002
- DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, 46, 137-149.
- Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, 27, 494-507. doi:10.1177/0734282909333179
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What no child left behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104, 439-451.  
doi:10.1037/a0026280
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91, 40-57. doi:10.1037/0021-9010.91.1.40
- Eurostat. (2014). *The EU in the world—population*. Retrieved from  
[http://ec.europa.eu/eurostat/statistics-explained/index.php/The\\_EU\\_in\\_the\\_world\\_-](http://ec.europa.eu/eurostat/statistics-explained/index.php/The_EU_in_the_world_-)

\_population#Migration\_and\_asylum

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: Wiley.

Flere, S., Krajnc, M. T., Klanjšek, R., Musil, B., & Kirbiš, A. (2010). Cultural capital and intellectual ability as predictors of scholastic achievement: A study of Slovenian secondary school students. *British Journal of Sociology of Education*, 31, 47-58. doi:10.1080/01425690903385428

Ganzach, Y., & Fried, I. (2012). The role of intelligence in the formation of well-being: From job rewards to job satisfaction. *Intelligence*, 40, 333-342. doi:10.1016/j.intell.2012.03.004

Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC-III. In J. Georgas, F. J. R. van de Vijver, L. G. Weiss, & D. H. Saklofske (Eds.), *Practical resources for the mental health professional: Culture and children's intelligence. Cross-cultural analysis of the WISC-III* (pp. 277-313). Amsterdam, Netherland: Elsevier. doi:10.1016/B978-012280055-9/50021-7

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, 40, 103-114. doi:10.1177/00224669060400020101

Goldstein, S., Princiotta, D., & Naglieri, J. A. (Eds.). (2015). *Handbook of intelligence. Evolutionary theory, historical perspective, and current concepts*. New York, NY: Springer. doi:10.1007/978-1-4939-1562-0

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, 24, 13-23. doi:10.1016/S0160-2896(97)90011-8

Gottfredson, L., & Deary, I. J. (2004). Intelligence predicts health and longevity: But why?

- Current Directions in Psychological Science*, 13, 1-4. doi:10.1111/j.0963-7214.2004.01301001.x
- Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology*, 50, 183-195. doi: 10.1037/a0016641
- Grob, A., Gygi, J. T., Hagmann-von Arx, P. (in preparation). *The Stanford-Binet Intelligence Scales, Fifth Edition (SB5) – German Adaptation*. Bern, Switzerland: Hogrefe.
- Grob, A., & Hagmann-von Arx, P. (in preparation). *Intelligence and Development Scales for Children and Adolescents (IDS-2)*. Bern, Switzerland: Hogrefe.
- Grob, A., Meyer, C. S., & Hagmann-von Arx, P. (2009). *Intelligence and Development Scales (IDS)*. Bern, Switzerland: Hans Huber.
- Grob, A., Meyer, C. S., & Hagmann-von Arx, P. (2013). *Intelligence and Development Scales (IDS)* (2nd ed.). Bern, Switzerland: Hans Huber.
- Grob, A., Reiman, G., Gut, J., & Frischknecht, M.-C. (2013). *Intelligence and Development Scales–Preschool (IDS-P)*. Bern, Switzerland: Hans Huber.
- Gut, J., Reimann, G., & Grob, A. (2012). Kognitive, sprachliche, mathematische und sozial-emotionale Kompetenzen als Prädiktoren späterer schulischer Leistungen. Können die Leistungen eines Kindes in den IDS dessen Schulleistungen drei Jahre später vorhersagen? [Cognitive, language, mathematical, and socioemotional skills as predictors of subsequent school performance: Is a child's performance on the Intelligence Development Scales predictive of his or her school performance three years later?]. *Zeitschrift für Pädagogische Psychologie*, 26, 213-200. doi:10.1024/1010-0652/a000070
- Gut, J., Reimann, G., & Grob, A. (2013). A contextualized view on long-term predictors of academic performance. *Journal of Educational Psychology*, 105, 436-443. doi:10.1037/a0031503
- Hagmann-von Arx, P., Gauck, L., & Grob, A. (2015). Intelligenzdiagnostik [Intelligence assessment]. In K. Seifried, S. Drewes, & M. Hasselhorn (Eds.), *Handbuch*

- Schulpsychologie* [Handbook of school psychology] (pp. 128-139). Stuttgart, Germany: Kohlhammer.
- Hagmann-von Arx, P., & Grob, A. (2014). *Reynolds Intellectual Assessment Scales (RIAS)*. Bern, Switzerland: Hans Huber.
- Hagmann-von Arx, P., Lemola, S., & Grob, A. (2016). Does IQ = IQ? Comparability of intelligence test scores in typically developing children. *Assessment*. doi:10.1177/1073191116662911
- Hagmann-von Arx, P., Petermann, F., & Grob, A. (2013). Konvergente und diskriminante Validität der WISC-IV und der Intelligence and Development Scales (IDS) bei Kindern mit Migrationshintergrund [Convergent and discriminant validity of the WISC-IV and the Intelligence and Development Scales (IDS) in children with a migration background]. *Diagnostica*, 59, 170-182. doi:10.1026/0012-1924/a000091
- Harris, J. G., Muñoz, M. R., & Llorente, A. M. (2008). Considerations in using the WISC-IV with Hispanic children. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention* (pp. 409-442). San Diego, CA: Elsevier.
- Hartmann, P., & Andresen, J. (2011). *Dokumentation af den danske standardisering af RIAS*. Virum, Denmark: Hogrefe Psykologisk.
- Hessels, M. G. (1997). Low IQ but high learning potential: Why Zeyneb and Moussa do not belong in. *Educational and Child Psychology*, 14, 121-136.
- Hofstede, G. (2001). *Culture's consequences comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual* (pp. 197-232). Chicago, IL: Riverside.

- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.  
doi:10.1037/h0023816
- Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: gf-gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York, NY: Guilford Press.
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The big five revisited. *Journal of Applied Psychology*, 85, 869-879. doi:10.1037/0021-9010.85.6.869.
- International Labour Organization. (2008). *International Standard Classification of Occupations 2008 (ISCO-08). Structure, group definitions and correspondence tables*. Geneva, Switzerland: International Labour Office. Retrieved from [http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms\\_172572.pdf](http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf)
- Irwin, J. K., Joschko, M., & Kerns, K. A. (2014). Confirmatory factor analysis of the Reynolds Intellectual Assessment Scales (RIAS) in Canadian children. *The Clinical Neuropsychologist*, 28, 1258-1277. doi:10.1080/13854046.2014.975843
- Judge, T. A., Heller, D., & Mount, M. K. (2002). Five-factor model of personality and job satisfaction: A meta-analysis. *Journal of Applied Psychology*, 87, 530-541.  
doi:10.1037/0021-9010.87.3.530
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, 98, 875-925. doi:10.1037/a0033901
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44, 486-507.  
doi:10.1177/0049124114543236

- Klingner, J. K., Blanchett, W. J., & Harry, B. (2007). Race, culture, and developmental disabilities. In S. L. Odom, R. H. Horner, & M. E. Snell (Eds.). *Handbook of developmental disabilities* (pp. 55-75). New York, NY: Guilford Press.
- Kramer, J. (1972). *Intelligenztest: Mit einer Einführung in Theorie und Praxis der Intelligenzprüfung* (4<sup>th</sup> edition). Solothurn, Switzerland: Antonius.
- Lahaie, C. (2008). School readiness of children of immigrants: Does parental involvement play a role? *Social Science Quarterly*, 89, 684-705. doi:10.1111/j.1540-6237.2008.00554.x
- Lounsbury, J. W., Gibson, L. W., Steel, R. P., Sundstrom, E. D., & Loveland J. L. (2004). An investigation of intelligence and personality in relation to career satisfaction. *Personality and Individual Differences*, 37, 181-189. doi:10.1016/J.Paid.2003.08.010
- Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford, England: Oxford University Press.
- Mayes, S. D., & Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children-Third and -Fourth Edition predictors of academic achievement in children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly*, 22, 234-249. doi:10.1037/1045-3830.22.2.234
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136-181). New York, NY: Guilford Press.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568-592. doi:10.1037/0021-9010.93.3.568
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 48, 525-543. doi:10.1007/BF02294825

- Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44, 69-77. doi:10.1097/01.mlr.0000245438.73837.89
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749. doi:10.1037/0003-066X.50.9.741
- Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3, 111-130. doi:10.21500/20112084.857.
- Mleko, A. L., & Burns, T. G. (2005). Test review. *Applied Neuropsychology*, 12, 179-180. doi:10.1207/S15324826AN1203\_8
- Moffitt, T. E., Caspi, A., Harkness, A. R., & Silva, P. A. (1993). The natural history of change to intellectual performance: Who changes? How much? Is it meaningful? *Journal of Child Psychology and Psychiatry*, 34, 455-506. doi:10.1111/j.1469-7610.1993.tb01031.x
- Moosbrugger, H., & Kelava, A. (2012) *Testtheorie und Fragebogenkonstruktion* [Test theory and questionnaire construction]. Heidelberg, Germany: Springer.
- Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment*, 24, 129-140. doi:10.1037/a0024878
- Nelson, J. M., Canivez, G. L., Lindstrom, W., & Hatt, C. V. (2007). Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *Journal of School Psychology*, 45, 439-456. doi:10.1016/j.jsp.2007.03.003
- Neuberger, O., & Allerbeck, M. (1978). *Messung und Analyse von Arbeitszufriedenheit* [Measurement and analysis of satisfaction with work]. Bern, Switzerland: Hans Huber.
- Neukrug, E. S., & Fawcett, R. C. (2015). *The essentials of testing and assessment: A practical guide for counselors, social workers, and psychologists* (3rd ed.). Stamford, CT: Cengage Learning.

- Ng, T. W., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personality Psychology, 58*, 367-408. doi:10.1111/j.1744-6570.2005.00515.x
- Oberle, E., Schonert-Reichl, K. A., Hertzman, C., & Zumbo, B. D. (2014). Social-emotional competencies make the grade: Predicting academic success in early adolescence. *Journal of Applied Developmental Psychology, 35*, 138-147. doi:10.1016/j.appdev.2014.02.004
- Ostendorf, F., & Angleitner, A. (2004). *Neo Persönlichkeitsinventar nach Costa und McCrae* [NEO Personality Inventory according to Costa and McCrae Revised]. Göttingen, Germany: Hogrefe.
- Petermann, F., & Petermann, U. (2011). *WISC-IV: Wechsler Intelligence Scale for Children—Fourth Edition* (German version). Frankfurt, Germany: Pearson Assessment.
- Petermann, F., & Renner, G. (2010). *Fallbuch SON-R 2<sup>1/2</sup>–7* [Casebook SON-R 2<sup>1/2</sup>–7]. Göttingen, Germany: Hogrefe.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322-338. doi:10.1037/a0014996
- Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and individual differences, 20*, 110-122. doi:10.1016/j.lindif.2009.10.005
- Resing, W. C., Tunteler, E., de Jong, F. M., & Bosma, T. (2009). Dynamic testing in indigenous and ethnic minority children. *Learning and Individual Differences, 19*, 445-450. doi:10.1016/j.lindif.2009.03.006
- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*, 700-712. doi:10.1037/a0027268
- Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales: Professional Manual*. Lutz, FL: PAR.



- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2, 313-345. doi:10.1111/j.1745-6916.2007.00047.x
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118-137. doi:10.1016/j.intell.2015.09.002
- Roid, G. (2003a). *Stanford-Binet Intelligence Scales: Fifth Edition*. Itasca, IL: Riverside Publishing.
- Roid, G. (2003b). *Stanford-Binet Intelligence Scales: Fifth Edition, Technical Manual*. Itasca, IL: Riverside Publishing.
- Rost, D. H. (2009). *Intelligenz: Fakten und Mythen* [Intelligence: Facts and myths]. Weinheim, Germany: Beltz.
- Ruhlen, M. (1991). *A guide to the world's languages: Volume I, classification*. Stanford, CA: Stanford University Press.
- Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspectives in Psychological Science*, 9, 538-551. doi:10.1177/1745691614543972
- Santamaría, P., & Fernández Pinto, I. (2009). *RIAS Escalas de Inteligencia de Reynolds y RIST Test de Inteligencia Breve de Reynolds Manual. Adaptación española*. Madrid, Spain: TEA Ediciones.
- Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: A commentary on four recent articles. *Perspectives of Psychological Science*, 9, 211–218. doi:10.1177/1745691613518074

- Schmidt, F. L., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274. doi:10.1037/0033-2909.124.2.262
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162-173. doi:10.1037/0022-3514.86.1.162
- Schneider, W. J., & Flanagan, D. P. (2015). The relationship between theories of intelligence and intelligence tests. In S. Goldstein, D. Princiotta, & J. A. Naglieri, (Eds.), *Handbook of intelligence. Evolutionary theory, historical perspective, and current concepts* (pp. 317-340). New York, NY: Springer. doi:10.1007/978-1-4939-1562-0\_21
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York, NY: Guilford Press.
- Schölmerich, A., & Leyendecker, B. (2009). Psychologische Diagnostik bei Kindern aus zugewanderten Familien. [Psychological assessment in children from families with a migration background]. In D. Irblich & G. Renner (Eds.), *Diagnostik in der Klinischen Kinderpsychologie. Die ersten sieben Lebensjahre* [Diagnostics in clinical child psychology. The first seven years of life] (pp. 430-442). Göttingen, Germany: Hogrefe.
- Snijders-Oomen, N. (1943). *Intelligentieonderzoek van doofstommekinderen*. Nijmegen, Netherland: Berhout.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, 15, 201-292. doi:10.2307/1412107
- Statistisches Bundesamt. (2015). *Bevölkerungsstand. [Population.]*. Retrieved from <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsstand.html;jsessionid=D9339CCA1586CB4B47CA4D661CFD1EE8.cae1>

- Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*, 80-90.  
doi:10.1016/j.lindif.2008.05.004
- Sternberg, R. J., Grigorenko, E., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly, 47*, 1-41. doi:10.1353/mpq.2001.0005
- Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence, 35*, 401-426. doi:10.1016/j.intell.2006.09.004
- Tellegen, P. J., Laros, J. A., & Petermann, F. (2007). *SON-R 2<sup>1/2</sup>-7: Nonverbaler Intelligenztest. Testmanual mit deutscher Normierung und Validierung*. Göttingen, Germany: Hogrefe.
- Tellegen, P. J., Laros, J. A., & Petermann, F. (2012). *Snijders-Oomen Nonverbal Intelligence Test (SON-R 6–40)*. Göttingen, Germany: Hogrefe.
- Terman, L. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale*. Boston, MA: Houghton Mifflin.
- Terman, L., & Merrill, M. (1937). *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence*. Boston, MA: Houghton Mifflin.
- Terman, L., & Merrill, M. (1960). *Stanford-Binet Intelligence Scale: Manual for the third revision form L-M*. Boston, MA: Houghton Mifflin.
- Thorndike, R., Hagen, E., & Sattler, J. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition. Technical manual*. Rolling Meadows, IL: Riverside Publishing.
- Turney, K., & Kao, G. (2009). Barriers to school involvement: Are immigrant parents disadvantaged?. *The Journal of Educational Research, 102*, 257-271.  
doi:10.3200/JOER.102.4.257-271
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly

- performance relates to early mathematical skills. *Child Development*, 85, 1062-1076.  
doi:10.1111/cdev.12165.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101, 817-835. doi: 10.1037/a0016127
- Wasserman, J. D., & Tulskey, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 3-38). New York, NY: Guilford Press.
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, 15, 465-479. doi:10.1037/h0088802
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Wood.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) administration and scoring manual*. San Antonio, TX: Pearson.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)*. Bloomington, MN: Pearson.
- Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition (WPPSI-IV)*. Bloomington, MN: Pearson.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V)*. Bloomington, MN: Pearson.
- Weiss, L. G., Harris, J. G., Prifitera, A., Courville, T., Rolfhus, E., Saklofske, D. H., & Holdnack, J. A. (2006). WISC-IV interpretation in societal context. In L. G. Weiss, D. H. Saklofske, A. Prifitera, & J. A. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 1–57). San Diego, CA: Elsevier.
- Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, 62, 1066-1078. doi:10.1111/1467-8624.ep9112161646

- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association. doi:10.1037/10222-009
- Willis, J. O., Dumont, R., & Kaufman, A. S. (2011). Factor-analytic models of intelligence. In R. J. Sternberg, & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 39-57). New York, NY: Cambridge University Press.  
doi:10.1017/CBO9780511977244.004
- Yoakum, C. S., & Yerkes, R. M. (1920). *Army Mental Tests*. New York, NY: Henry Holt.  
Retrieved from  
<https://babel.hathitrust.org/cgi/pt?id=hvd.32044029556826;view=1up;seq=16>

**APPENDIX A: Study 1**

Gygi, J. T., Ledermann, T., Grob, A., & Hagmann-von Arx, P. The Reynolds Intellectual Assessment Scales (RIAS): Measurement invariance across four language groups.

*Manuscript submitted for publication to European Journal of Psychological Assessment.*

Draft December 1, 2016

Running Head: INVARIANCE OF THE RIAS

1

The Reynolds Intellectual Assessment Scales (RIAS): Measurement invariance across four  
language groups

Jasmin T. Gygi<sup>1</sup>, Thomas Ledermann<sup>2</sup>, Alexander Grob<sup>1</sup>, and Priska Hagmann-von Arx<sup>1</sup>

<sup>1</sup>University of Basel, <sup>2</sup>Utah State University

Acknowledgments:

We thank Hogrefe Psychologisk Forlag A/S, Virum, Denmark for providing us the data on the Danish RIAS standardization sample, and TEA Ediciones, Madrid for providing us the data on the Spanish RIAS standardization sample.

Special thanks to Cecil R. Reynolds for helpful comments.

Corresponding author:

Jasmin T. Gygi, Department of Psychology, University of Basel, Missionsstrasse 62, 4055 Basel, Switzerland. E-mail: [jasmin.gygi@unibas.ch](mailto:jasmin.gygi@unibas.ch); phone: 0041-61-267-06 21

## Abstract

The Reynolds Intellectual Assessment Scales (RIAS) is an intelligence test for individuals aged 3 to above 90 years developed in the United States. The RIAS measures general intelligence and its two main components, verbal and nonverbal intelligence, each comprising two subtests. Memory can be assessed through two additional subtests. The test has recently been standardized in other language groups, including the Danish version standardized in Denmark, the German version standardized in Switzerland and Germany, and the Spanish version standardized in Spain. However, it is unknown whether measurement invariance of the RIAS intelligence factor structure holds across the English, Danish, German, and Spanish version. This study examined measurement invariance of a single-factor structure, measuring general intelligence, and a two-factor structure, measuring verbal and nonverbal intelligence, across the four language groups. Standardization samples of the English ( $n = 2,438$ ), Danish ( $n = 983$ ), German ( $n = 2,103$ ), and Spanish ( $n = 1,933$ ) RIAS versions were analyzed. A series of confirmatory factor analysis (CFA) models were tested supporting the single-factor and two-factor intelligence structure for each language group. A multiple-group CFA supported scalar invariance across groups, indicating an identical factor structure across individuals from the United States, Denmark, Switzerland and Germany, and Spain.

*Keywords:* intelligence, Reynolds Intellectual Assessment Scales (RIAS), cross-cultural, measurement invariance.



## INVARIANCE OF THE RIAS

3

The Reynolds Intellectual Assessment Scales (RIAS): Measurement invariance across four language groups

Intelligence is one of the most often studied and measured psychological constructs (Goldstein, Princiotta, & Naglieri, 2015). The Reynolds Intellectual Assessment Scales (RIAS) is an individually administered intelligence test with a supplemental measure of memory appropriate for individuals aged 3 to above 90 years developed in the United States by Reynolds and Kamphaus (2003). Over the last decade, several studies have repeatedly demonstrated the convergent, discriminant, and predictive validity of the English RIAS version (Beaujean, Firmin, Michonski, Berry, & Johnson, 2010; Edwards & Paulin, 2007; Krach, Loe, Jones, & Farrally, 2009; Nelson & Canivez, 2012). Moreover, the RIAS intelligence measure has been evaluated to be independent of motor coordination, visual-motor speed, and reading skills, and has been found to be user friendly in terms of its administration, scoring, and interpretation (Andrews, 2007; Dombrowski & Mrazik, 2008; Elliot, 2004). Recently, the RIAS was successfully adapted to three further language groups: The Danish version (Hartmann & Andresen, 2011) standardized in Denmark, the German version (Hagmann-von Arx & Grob, 2014) standardized in Switzerland and Germany, and the Spanish version (Santamaría & Fernández Pinto, 2009) standardized in Spain.

The theoretically proposed factor structure of the RIAS is based on Carroll's (1993) three-stratum theory. The RIAS consists of four intelligence subtests on Stratum 1 and two factors—Verbal Intelligence Index (VIX) and Nonverbal Intelligence Index (NIX)—on Stratum 2. These factors serve as indicators of crystallized and fluid intelligence, which in turn are combined to a Composite Intelligence Index (CIX) on Stratum 3, which reflects general intelligence, *g*. There is also a Composite Memory Index (CMX) on Stratum 2, which is composed of scores from two supplemental subtests and is not integrated into the measure of general intelligence. Based on the theoretical assumptions, Reynolds and Kamphaus (2003) evaluated a series of alternative factor structures for the RIAS and posit a structure with four subtests measuring two factors, verbal and nonverbal, (Figure 1a). One alternative structure is

---

Figure 1

---

Analyses assessing the factor structure of psychological tests have been found to be important in ensuring that test scores can be interpreted according to the posited test structure (e.g., Widaman & Reise, 1997). The RIAS intelligence factor structure proposed by Reynolds and Kamphaus (2003; Figure 1a) has been repeatedly supported in three other language groups besides English, including Danish, German, and Spanish, by using standardization samples (Hagmann-von Arx & Grob, 2014; Hartmann & Andresen, 2011; Reynolds & Kamphaus, 2003; Santamaría & Fernández Pinto, 2009). This factor structure was further supported in an independent study with a referred U.S. student sample (Beaujean, McGlaughlin, & Margulies, 2009). However, studies have also found evidence for a single-factor structure (Figure 1b) for the English RIAS standardization sample (Dombrowski, Watkins, & Brogan, 2009) and for referred samples (Beaujean & McGlaughlin, 2014; Nelson & Canivez, 2012; Nelson, Canivez, Lindstrom, & Hatt, 2007).

While the theoretically proposed RIAS intelligence factor structure was supported in each individual language group providing evidence for configural invariance of the RIAS (Meredith, 1993), it is unknown whether higher levels of factorial (measurement) invariance hold across the four language groups. The higher levels of measurement invariance are (e.g., Meredith, 1993; Widaman & Reise, 1997): metric invariance (invariant factor loadings across groups), scalar invariance (invariant factor loadings and intercepts across groups), and residual invariance (invariant factor loadings, intercepts, and residual variances across groups). Configural invariance indicates that the factor structure is the same across groups but does not guarantee that the observed variables measure the latent construct in the same way. Metric invariance implies that the observed variables are related to the latent variable equivalently across groups. This level of invariance allows a comparison of the groups in terms of path coefficients and covariances between observed and latent variables (Chen,

## INVARIANCE OF THE RIAS

5

(Widaman & Reise, 1997). This level of invariance allows practitioners to compare results of different language versions and assess individuals with a migration background in their native language when they have insufficient language skills in the country of resettlement (Sattler, 2001), which has become increasingly important as migration has increased worldwide in recent decades (Eurostat, 2014; U.S. Department of Homeland Security, 2013). Residual invariance allows a comparison of residual variances across groups (Chen, et al., 2005).

The goal of the present study was to assess measurement invariance of the English, Danish, German, and Spanish version of the RIAS. First, we conducted single-group confirmatory factor analysis (CFA) for each language group. Second, we ran a multiple-group CFA to assess measurement invariance across all groups simultaneously as well as in comparison of each group with the English version of the RIAS as reference group.

**Methods****Participants**

The study included a total of 7,457 individuals from the RIAS standardization samples. The English standardization sample includes 2,438 individuals from the United States, the Danish standardization sample includes 983 individuals from Denmark, the German standardization sample includes 2,103 individuals from Switzerland and Germany, and the Spanish standardization sample includes 1,933 individuals from Spain. All standardization samples are equally distributed with respect to sex (50% female) and are meant to match the census data on educational attainment of the respective countries. The English and Spanish RIAS versions are standardized for individuals aged 3 to 94 years; the German and Danish versions are standardized for individuals aged 3 to 99 years. Demographic characteristics are provided in detail in the Technical Manuals (Hagmann-von Arx & Grob, 2014; Hartmann & Andresen, 2011; Reynolds & Kamphaus, 2003; Santamaría & Fernández Pinto, 2009).

**Measures**

The RIAS is an individually administered intelligence test, developed in the United States (Reynolds & Kamphaus, 2003) and recently adapted to Danish (Hartmann & Andresen,

## INVARIANCE OF THE RIAS

6

specifically to the American culture were modified so that they were appropriate for the culture of the respective language group (e.g., the American Newspaper machine was replaced by a machine used in European countries). In addition, the items were empirically pretested and the order of the items was adjusted to ensure an order of ascending difficulty (e.g., questions regarding baseball were easier for Americans while questions regarding soccer were easier for Europeans). Finally, the adaptations were standardized using a representative sample of the respective language group.

The RIAS yields CIX (Composite Intelligence Index), VIX (Verbal Intelligence Index), and NIX (Nonverbal Intelligence Index). VIX and NIX comprise two subtests each. The verbal subtests are Guess What (GWH; identifying an object or a concept through the use of verbally presented clues) and Verbal Reasoning (VRZ; completing verbal analogies). The nonverbal subtests are Odd-Item Out (OIO; identifying a picture that does not go with the others) and What's Missing? (WHM; identifying the missing element of a presented picture). The CIX is calculated from the sum of the *T* scores ( $M = 50$ ;  $SD = 10$ ) of the four intelligence subtests. The RIAS also includes a conormed, supplemental Composite Memory Index (CMX) comprising two subtests, Verbal Memory (VRM; reproducing verbally presented sentences and short stories), and Nonverbal Memory (NVM; recognizing visually presented objects). In line with suggestions of others (e.g., Nelson & Canivez, 2012), the memory subtests were not included in the analysis. Reliability for the subtests of each of the four RIAS language versions is high with Cronbach's alphas ranging from .81 to .97.

**Data Analysis**

SPSS AMOS 20 (Arbuckle, 2011) was used to assess measurement invariance across the four language groups. Correlation matrices including variances and means of the subtest *T* scores ( $M = 50$ ;  $SD = 10$ ) were used as data input (see Online Supplemental Table 1). First, we conducted single-group CFAs, as recommended by Meade, Johnson, and Braddy (2008), to test both factor structures as shown in Figure 1. Second, we conducted a multiple-group CFA across the four language groups. We first analyzed all four groups simultaneously and

## INVARIANCE OF THE RIAS

7

factor structure, the intercepts of GWH and OIO were fixed to zero across all groups and for the single-factor structure, the intercept of the subtest GWH was fixed to zero in all groups. In both models, the factor loading of GWH was set to 1. For the two-factor structure, the factor loading of OIO was fixed to 1 in addition to the factor loading of GWH. Finally, for the single-factor structure the residuals of GWH and VRZ as well as the residuals of OIO and WHM were allowed to covary (Beaujean & McGlaughlin, 2014). These covariances were set to be equal across groups (Nelson et al., 2007). Next, metric invariance (Model 2), scalar invariance (Model 3), and residual invariance (Model 4) were assessed.

We relied on the following goodness-of-fit indices and criteria to assess the model fit: Comparative Fit Index (CFI) of  $\geq .95$ , McDonald's Noncentrality Index (Mc) of  $\geq .90$  and root mean square error of approximation (RMSEA) of  $\leq .06$  (Hu & Bentler, 1999). Because of the large sample size, we did not consider chi-square test statistics (Meade, et al., 2008) but rather used the test of small difference in fit to compare the nested models (MacCallum, Browne, & Cai, 2006). Additionally, we calculated the  $\Delta CFI$  and  $\Delta Mc$  and considered values of  $\leq .002$  for  $\Delta CFI$  and  $\leq .005$  for  $\Delta Mc$  as insignificant reduction in model fit, as suggested by Meade et al. (2008).

## Results

### Single-group CFA

For each language group, we tested the two-factor and single-factor structure with the four intelligence subtests as indicators. In accordance with Beaujean and McGlaughlin (2014), we allowed the residuals of GWH and VRZ as well as the residuals of OIO and WHM to covary in the single-factor structure, but constrained the covariances between these residuals to be equal (Nelson et al., 2007). We note that this setup results in a model that is empirically equivalent to the two-factor structure as well as to a bifactor structure with a general factor and two specific factors each measured by two indicators with fixed factor loadings (e.g., Canivez & Watkins, 2016; Reise, 2012). The fit estimates of the single-group CFA for the two-factor structure and the single-factor structure, which are empirically equivalent (Lee &

---

Table 1

---

**Multiple-group CFA**

Next, we assessed measurement invariance across all groups simultaneously using multiple-group CFA. Four models were tested implying configural, metric, scalar, and residual invariance, respectively. Results for the two-factor structure (Figure 1a) are shown in Table 2. Models 1, 2, and 3 showed a good model fit. Comparing Model 1 and 2 and Model 2 and 3, the test of small difference in fit,  $\Delta CFI$ , and  $\Delta Mc$  revealed no significant reduction in the model fit, indicating that scalar invariance holds across all four groups. Model 4 implying residual invariance yielded a poor fit, indicating that the variances of the error terms varied across groups.

---

Table 2

---

Results of the single-factor structure (Figure 1b) are given in Table 3 and are similar to those of the two-factor structure. Again, scalar invariance holds but not residual invariance.

---

Table 3

---

Finally, multiple-group CFAs were conducted comparing the original English version separately to each of the other language group (see supplementary results for details). Results for the two-factor structure indicated that the Danish and Spanish groups differed significantly from the English group in terms of residual invariance (Model 4). The German group showed invariance on all levels in comparison with the English group (see Online Supplemental Table 3). Results for the single-factor structure look similar: the Danish and Spanish groups showed noninvariance for the residual invariance (Model 4), whereas the German group showed

## INVARIANCE OF THE RIAS

9

Tables, the values of the CFI, Mc, and RMSEA are somewhat in favor of the two-factor structure.

**Discussion**

The RIAS is a widely used intelligence test developed and first published in the United States that has recently been made available in three additional language groups: Danish, German, and Spanish. This study assessed measurement invariance of the RIAS intelligence factor structure across these four language groups. We used CFA techniques and found evidence for the RIAS intelligence factor structure proposed by Reynolds and Kamphaus (2003) with four intelligence subtests and two factors (verbal and nonverbal). This finding is in line with the proposed factor structure based on Carroll's (1993) three-stratum theory and replicates findings of previous studies (Beaujean et al., 2009; Hagmann-von Arx & Grob, 2014; Hartmann & Andresen, 2011; Reynolds & Kamphaus, 2003; Santamaría & Fernández Pinto, 2009). Using single-group CFA, the two-factor structure measuring verbal and nonverbal intelligence is empirically equivalent to the single-factor structure with covariances between the two verbal tests and between the two nonverbal test. The single-factor structure is, in turn, equivalent to a bifactor structure measuring general intelligence and two specific factors. Thus, the single-group CFA is also in support of the single-factor structure, which is in line with other studies (Beaujean & McGlaughlin, 2014; Dombrowski et al., 2009; Nelson & Canivez, 2012; Nelson et al., 2007). Using multiple-group CFA, the fit measures used in the current study indicate that the two-factor structure fits the data slightly better than the single-factor structure, a finding which is in line with the notion that the two-factor structure is theoretically more established (Carroll, 1993; Reynolds & Kamphaus, 2003). We further note that high correlations were found between the verbal and nonverbal factors, which is in support of either a higher-order factor, *g*, or a bifactor structure (Canivez, in press; Reise, 2012). However, because the RIAS only consists of four intelligence subtests and thus only comprises two indicators per factor, all these models are statistically indistinguishable (Brown, 2006; Lee & Hersberger, 1990; Rindskopf & Rose, 1988).

## INVARIANCE OF THE RIAS

10

variances varied across groups), which may be due to varying reliabilities of the scales across groups (DeShon, 2004). However, residual invariance has been considered overly restrictive and therefore may be negligible (e.g., Widaman & Reise, 1997).

Finally, separate multiple-group CFAs with the original English version as reference group revealed that the German group showed invariance in all models for the two-factor and single-factor structure, whereas the Danish group and the Spanish group showed noninvariance in the residual invariance. This indicates that the RIAS intelligence structure and the scores in these groups are mostly comparable. However, the Danish and Spanish groups showed noninvariance in residual variances, indicating differing reliabilities of the scales (DeShon, 2004).

There are strengths and limitations. It is a strength that the RIAS intelligence factor structure could be analyzed across all language groups in which the RIAS currently is available. Also, this study is based on impressive large samples. Another strength is the employment of confirmatory methods. It should be noted, though, that analyses were conducted using the standardized scores of the standardization samples as raw data were not made available by some publishers. The use of standardized scores, which expresses the extent to which a person deviates from the norm, is common and in line with other studies (Beaujean, Freeman, Youngstrom, & Carlson, 2012; Beaujean & McGlaughlin, 2014; Bialer, 1974; Naglieri, Taddei, & Williams, 2013; Nelson & Canivez, 2012). However, future studies may use raw scores, which would allow researchers to address additional questions, such as whether general populations differ in their factor means. Due to the wide age range of the RIAS, the raw scores are typically highly correlated with age (Reynolds & Kamphaus, 2003). Therefore, analyses using raw scores would require the assessment of age-homogeneous groups (cf. Bialer, 1974). A second limitation of the current study is the use of samples representing the general populations and it remains unknown whether measurement invariance holds for specific subpopulations. There is one study focusing on referred students, which found evidence that the English RIAS intelligence factor structure holds across samples



## INVARIANCE OF THE RIAS

11

with small degrees of freedom (c.f. Kenny, Kaniskan, & McCoach, 2015). However, the use of large sample sizes alleviates this problem at least for the RMSEA (Kenny et al., 2015). Finally, invariance was examined at the level of subtests assuming similar loadings of the items across the language groups. Future research may additionally analyze measurement invariance at the item-level using raw scores, which were also not made available to other researchers by some test publishers (c.f. Beaujean et al., 2009).

In conclusion, our study strongly supports the theoretically proposed RIAS intelligence factor structures with a verbal and nonverbal intelligence factor, each comprising two subtests, and a general intelligence factor comprising four subtests, indicating that these RIAS indices can be used to gather valid diagnostic information. Further, the RIAS factor structures showed measurement invariance across all four language groups, indicating that the RIAS test structure is comparable across individuals from the United States, Denmark, Switzerland and Germany, and Spain.

## References

- Andrews, J. J. W. (2007). RIAS: Reynolds Intellectual Assessment Scales. *Journal of Psychoeducational Assessment, 25*, 402-408.
- Arbuckle, J. L. (2011). *IBM SPSS Amos 20 user's guide*. Armonk, NY: IBM.
- Beaujean, A. A., Firmin, M. W., Michonski, J. D., Berry, T., & Johnson, C. (2010). A multitrait–multimethod examination of the Reynolds Intellectual Assessment Scales in a college sample. *Assessment, 17*, 347-360. doi:10.1177/1073191109356865
- Beaujean, A.A., Freeman, M.J., Youngstrom, E., & Carlson, G. (2012). The structure of Cognitive Abilities in youths with manic symptoms: A factorial invariance study. *Assessment, 19*, 462-471. doi:10.1177/1073191111399037
- Beaujean, A. A., McGlaughlin, S. M., & Margulies, A. S. (2009). Factorial validity of the Reynolds Intellectual Assessment Scales for referred students. *Psychology in the Schools, 46*, 932-950. doi:10.1002/pits.20435
- Beaujean, A. A., & McGlaughlin, S. M. (2014). Invariance in the Reynolds Intellectual Assessment Scales for Black and White referred students. *Psychological Assessment, 26*, 1394. doi: [10.1037/pas0000029](https://doi.org/10.1037/pas0000029)
- Bialer, I. R. V. (1974). Raw versus standardized intelligence test scores. *The British Journal of Mental Subnormality, 20*, 93-96. doi:10.1179/bjms.1974.017
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Canivez, G. L., & Watkins, M. W. (2016). Review of the Wechsler Intelligence Scale for Children–Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Authors), *Intelligent testing with the WISC–V* (pp. 683–702). Hoboken, NJ: Wiley.
- Canivez, G. L. (in press). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards*

## INVARIANCE OF THE RIAS

13

- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling*, 12, 471-492.
- DeShon, R. P. (2004). Measures are not invariant across groups without error variance homogeneity. *Psychology Science*, 46, 137-149.
- Dombrowski, S. C., & Mrazik, M. (2008). Test review: Reynolds Intellectual Assessment Scales. *Canadian Journal of School Psychology*, 23, 223-230.
- Dombrowski, S. C., Watkins, M. W., & Brogan, M. J. (2009). An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *Journal of Psychoeducational Assessment*, 27, 494-507. doi:10.1177/0734282909333179
- Edwards, O. W., & Paulin, R. V. (2007). Referred students' performance on the Reynolds Intellectual Assessment Scales and the Wechsler Intelligence Scale for Children. *Journal of Psychoeducational Assessment*, 25, 334-340. doi:10.1177/0734282907300453
- Elliot, R. W. (2004). Test review: Reynolds Intellectual Assessment Scales. *Archives of Clinical Neuropsychology*, 19, 325-328.
- Eurostat. (2014). The EU in the world—population. Retrieved from [http://ec.europa.eu/eurostat/statistics-explained/index.php/The\\_EU\\_in\\_the\\_world\\_-\\_population#Migration\\_and\\_asylum](http://ec.europa.eu/eurostat/statistics-explained/index.php/The_EU_in_the_world_-_population#Migration_and_asylum)
- Goldstein, S., Princiotta, D., & Naglieri, J. A. (Eds.). (2015). *Handbook of intelligence. Evolutionary theory, historical perspective, and current concepts*. doi:10.1007/978-1-4939-1562-0
- Hagmann-von Arx, P., & Grob, A. (2014). *Reynolds Intellectual Assessment Scales (RIAS)*. Bern: Hans Huber.
- Hartmann, P., & Andresen, J. (2011). *Dokumentation af den danske standardisering af RIAS*. Virum, Denmark: Hogrefe Psykologisk.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55. doi:10.1080/10705519909540118

## INVARIANCE OF THE RIAS

14

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling. Third Edition.*

New York, NY: The Guilford.

Krach, S. K., Loe, S. A., Jones, W. P., & Farrally, A. (2009). Convergent validity of the Reynolds Intellectual Assessment Scales (RIAS) using the Woodcock–Johnson Tests of Cognitive Ability, (WJ-III) with university students. *Journal of Psychoeducational Assessment*, 27, 355-365. doi:10.1177/0734282909331749

Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, 25, 313-334. doi:10.1207/s15327906mbr2503\_4

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*, 11, 19-35. doi:10.1037/1082-989X.11.1.19

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568-592. doi:10.1037/0021-9010.93.3.568

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 48, 525-543. doi:10.1007/BF02294825

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3, 111-130.

Naglieri, J.A., Taddei, S., & Williams, K.M. (2013). Multigroup Confirmatory Factor Analysis of U.S. and Italian Children's Performance on the PASS Theory of Intelligence as Measured by the Cognitive Assessment System. *Psychological Assessment*, 25, 157-166. doi: 10.1037/a0029828

Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment*, 24, 129-140. doi:10.1037/a0024878

## INVARIANCE OF THE RIAS

15

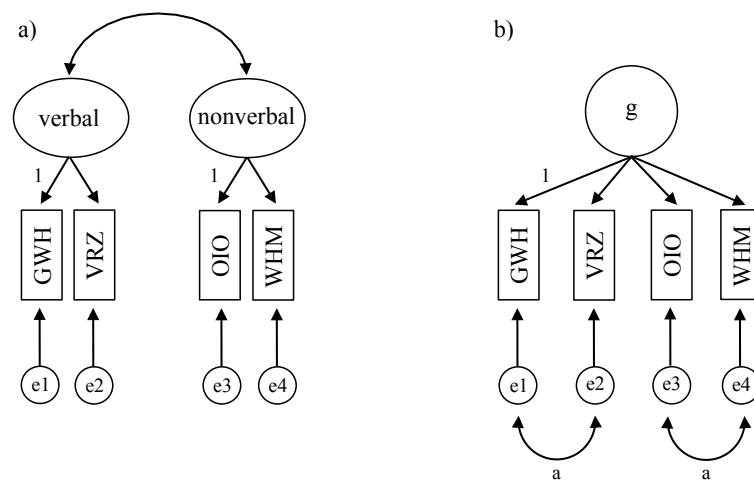
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. doi:10.1080/00273171.2012.715555
- Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales: Professional Manual*. Lutz, FL: PAR.
- Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23, 51-67. doi: 10.1207/s15327906mbr2301\_3
- Santamaría, P., & Fernández Pinto, I. (2009). *RIAS Escalas de Inteligencia de Reynolds y RIST Test de Inteligencia Breve de Reynolds Manual. Adaptación española*. Madrid, Spain: TEA Ediciones.
- Sattler, J. M. (2001). *Assessment of children. Cognitive applications*. La Mesa, CA: Sattler.
- Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-107. doi:10.1086/209528
- U. S. Department of Homeland Security. (2013). Yearbook of immigrant statistics. Retrieved from <http://www.dhs.gov/yearbook-immigration-statistics>.
- Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9, 486-492. doi: 10.1080/17405629.2012.686740
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. Bryant, M. Windle, & S. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association. doi:10.1037/10222-009

## INVARIANCE OF THE RIAS

16

*Figure 1.* The two-factor structure (a) and single-factor structure (b) of the RIAS. g = general intelligence; GWH = Guess What; VRZ = Verbal Reasoning; OIO = Odd-Item Out; WHM = What's Missing; VRM = Verbal Memory; NVM = Nonverbal Memory; e1-e4 = residuals; a = covariances of the residuals were set to be equal across groups.

Source of Figure 1: PLoS ONE, 11, e0166533.



**Figure 1.** The two-factor structure (a) and single-factor structure (b) of the RIAS.  $g$  = general intelligence; GWH = Guess What; VRZ = Verbal Reasoning; OIO = Odd-Item Out; WHM = What's Missing;  $e1$ - $e4$  = residuals;  $a$  = covariances of the residuals were set to be equal across groups.

Source of Figure 1: PLoS ONE, 11, e0166533.

## INVARIANCE OF THE RIAS

18

Table 1

*Fit Indices for Single-Group Confirmatory Factor Analyses Evaluating Two RIAS Intelligence Factor Structures in Four Language Groups.*

Structure	df	$\chi^2$	CFI	Mc	RMSEA	90% CI
a)						
English	1	0.526	1.000	1.000	<.001	[.000, .048]
Danish	1	1.047	1.000	1.000	.007	[.000, .085]
German	1	0.245	1.000	1.000	<.001	[.000, .046]
Spanish	1	2.866	.999	1.000	.031	[.000, .076]
b)						
English	1	0.526	1.000	1.000	<.001	[.000, .048]
Danish	1	1.047	1.000	1.000	.007	[.000, .085]
German	1	0.245	1.000	1.000	<.001	[.000, .046]
Spanish	1	2.866	.999	1.000	.031	[.000, .076]

*Note.*  $N_{\text{English}} = 2,438$ ,  $N_{\text{Danish}} = 983$ ,  $N_{\text{German}} = 2,103$ ,  $N_{\text{Spanish}} = 1,933$ . a) two-factor structure including four subtests, b) single-factor structure including four subtests. CFI = Comparative Fit Index, Mc = McDonald's Noncentrality Index; RMSEA = root-mean-square error of approximation.



Running Head: INVARIANCE OF THE RIAS

19

Table 2

*Fit Indices for Multiple-Group Confirmatory Factor Analysis Evaluating Measurement Invariance of the Two-Factor Structure Including the Four RIAS Intelligence Subtests Across Four Language Groups.*

Model	<i>df</i>	$\chi^2$	CFI	Mc	RMSEA	90% CI	$\Delta df$	$\Delta \chi^2$	$\Delta CFI$	$\Delta Mc$	Small diff <i>p</i>
1. Configural Invariance	4	4.684	1.000	.999	.005	[.000, .019]	–	–	–	–	–
2. Metric Invariance	10	22.778	.999	.999	.013	[.006, .020]	–	–	–	–	–
2 versus 1	–	–	–	–	–	–	6	18.094	.001	<.001	.148
3. Scalar Invariance	16	23.926	.999	.999	.008	[.000, .015]	–	–	–	–	–
3 versus 2	–	–	–	–	–	–	6	1.148	<.001	<.001	.979
4. Residual Invariance	28	1652.957	.879	.897	.088	[.085, .092]	–	–	–	–	–
4 versus 3	–	–	–	–	–	–	12	1629.031	.120	0.102	<.001

*Note.*  $N_{\text{total}} = 7,457$ ,  $N_{\text{English}} = 2,438$ ,  $N_{\text{Danish}} = 983$ ,  $N_{\text{German}} = 2,103$ ,  $N_{\text{Spanish}} = 1,933$ . CFI = Comparative Fit Index, Mc = McDonald's Noncentrality Index,

RMSEA = root-mean-square error of approximation.

## INVARIANCE OF THE RIAS

20

Table 3

*Fit Indices for Multiple-Group Confirmatory Factor Analysis Evaluating Measurement Invariance of the Single-Factor Structure Including the Four RIAS Intelligence Subtests Across Four Language Groups.*

Model	<i>df</i>	$\chi^2$	CFI	Mc	RMSEA	90% CI	$\Delta df$	$\Delta \chi^2$	$\Delta CFI$	$\Delta Mc$	Small diff <i>p</i>
1. Configural Invariance	7	109.424	.992	.993	.044	[.037, .052]	–	–	–	–	–
2. Metric Invariance	16	192.492	.987	.988	.038	[.034, .043]	–	–	–	–	–
2 versus 1	–	–	–	–	–	–	9	83.068	.005	.005	.005
3. Scalar Invariance	25	193.947	.987	.989	.030	[.026, .034]	–	–	–	–	–
3 versus 2	–	–	–	–	–	–	9	1.455	<.001	<.001	.997
4. Residual Invariance	37	2224.182	.837	.864	.089	[.086, .092]	–	–	–	–	–
4 versus 3	–	–	–	–	–	–	12	2030.235	.150	.125	<.001

*Note.*  $N_{total} = 7,457$ ,  $N_{English} = 2,438$ ,  $N_{Danish} = 983$ ,  $N_{German} = 2,103$ ,  $N_{Spanish} = 1,933$ . CFI = Comparative Fit Index, Mc = McDonald's Noncentrality Index,

RMSEA = root-mean-square error of approximation.

**APPENDIX B: Study 2**

Gygi, J. T., Fux, E., Grob, A., & Hagmann-von Arx, P. (2016). Measurement invariance and latent mean differences in the Reynolds Intellectual Assessment Scales (RIAS):

Does the German version of the RIAS allow a valid assessment of individuals with a migration background? *PLoS ONE*, *11*, e0166533.

## RESEARCH ARTICLE

# Measurement Invariance and Latent Mean Differences in the Reynolds Intellectual Assessment Scales (RIAS): Does the German Version of the RIAS Allow a Valid Assessment of Individuals with a Migration Background?

Jasmin T. Gygi\*, Elodie Fux, Alexander Grob, Priska Hagmann-von Arx

Department of Psychology, University of Basel, Basel, Switzerland

\* [jasmin.gygi@unibas.ch](mailto:jasmin.gygi@unibas.ch)



## OPEN ACCESS

**Citation:** Gygi JT, Fux E, Grob A, Hagmann-von Arx P (2016) Measurement Invariance and Latent Mean Differences in the Reynolds Intellectual Assessment Scales (RIAS): Does the German Version of the RIAS Allow a Valid Assessment of Individuals with a Migration Background? PLoS ONE 11(11): e0166533. doi:10.1371/journal.pone.0166533

**Editor:** Pietro Cipresso, IRCCS Istituto Auxologico Italiano, ITALY

**Received:** June 27, 2016

**Accepted:** October 31, 2016

**Published:** November 15, 2016

**Copyright:** © 2016 Gygi et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

This study examined measurement invariance and latent mean differences in the German version of the Reynolds Intellectual Assessment Scales (RIAS) for 316 individuals with a migration background (defined as speaking German as a second language) and 316 sex- and age-matched natives. The RIAS measures general intelligence (single-factor structure) and its two components, verbal and nonverbal intelligence (two-factor structure). Results of a multi-group confirmatory factor analysis showed scalar invariance for the two-factor and partial scalar invariance for the single-factor structure. We conclude that the two-factor structure of the RIAS is comparable across groups. Hence, verbal and nonverbal intelligence but not general intelligence should be considered when comparing RIAS test results of individuals with and without a migration background. Further, latent mean differences especially on the verbal, but also on the nonverbal intelligence index indicate language barriers for individuals with a migration background, as subtests corresponding to verbal intelligence require higher skills in German language. Moreover, cultural, environmental, and social factors that have to be taken into account when assessing individuals with a migration background are discussed.

## Introduction

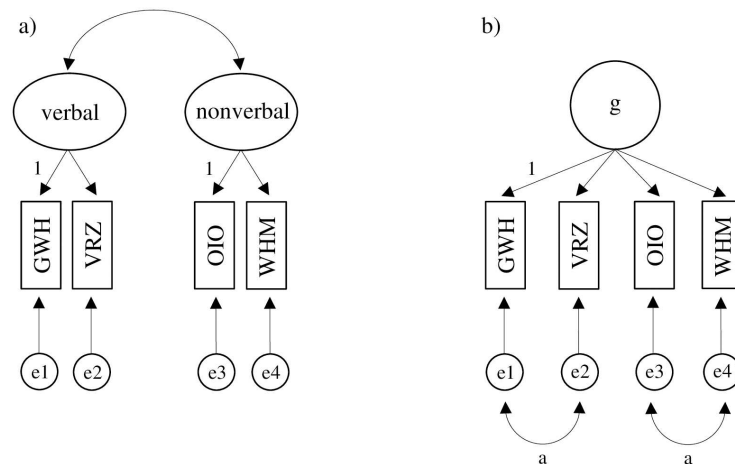
Migration has increased in recent decades in Europe [1]. For German-speaking countries, for example, the latest numbers reveal that 25% of the Swiss population [2] and 21% of the German population [3] have a migration background. The majority of immigrants originated from other European countries, including, for example, Italy and Turkey, while a smaller percentage originated from non-Western countries. The immigration of non-German-speaking individuals into Switzerland and Germany leads to more people seeking psychological assessment in a foreign-language environment in the respective countries.

Intelligence is one of the constructs most often studied in psychological practice [4]. Yet in individuals with a migration background, language barriers can lead to difficulties during assessment [5–7] such as misunderstanding of instructions or difficulties in articulating a verbal response. Thus, lower test scores in individuals with a migration background are particularly evident in tasks with high language requirements and their test performance increases with decreasing language dependence of the tasks [6–9]. Moreover, cultural and environmental factors may adversely affect test performance of individuals with a migration background [10]. For example, individuals with a migration background may be less familiar with the type of tasks used in performance assessment [11, 12]. In addition, many immigrant groups are confronted with negative achievement stereotypes, which may create increased performance pressure [13]. Furthermore, studies have shown that immigrant children's parents who have limited language skills in the dominant language are less involved in schooling, which in turn may negatively affect their children's performance [14, 15]. Consequently, in intelligence tests, individuals with a migration background may not be able to show their full potential, which can result in test scores that underestimate intelligence [5–7, 11, 12]. As high-stakes decisions are made on the basis of intelligence test scores, underestimated intelligence test scores may have negative consequences in, for example, education and employment. Regarding school-related decisions, underestimated intelligence test scores may lead to erroneous placement in special education services [16, 17]. In workplace situations, negative stereotypes may negatively influence performance ratings [18] and underestimated intelligence test scores may lead to not getting hired (e.g., [4]). Understanding how migration status influences test scores is therefore crucial to correctly interpreting the results.

To assess test score differences in individuals with and without a migration background, it is essential to examine measurement invariance prior to any interpretation of group differences [19]. Measurement invariance refers to the assumption of comparable relationships between items and their respective latent variables across groups [20]. When measurement invariance holds, researchers can validly compare statistical results, such as latent means, across these groups [21].

One currently used intelligence test is the Reynolds Intellectual Assessment Scales (RIAS) [22]. The RIAS is an individually administered test for individuals aged 3 to 90-plus years developed in the United States and recently adapted to different language groups including Danish [23], German [24], and Spanish [25]. The RIAS is designed to measure general intelligence and its two components, verbal and nonverbal intelligence. Additionally, the RIAS provides a measure of memory. Performance on the RIAS has been shown to be independent of motor coordination, visual-motor speed, and reading skills. Also, its administration, scoring, and interpretation have been described as user friendly [26–28]. However, we know of no study that has examined the RIAS intelligence factor structure across individuals with and without a migration background and whether these groups of individuals achieve comparable RIAS intelligence test scores.

The intelligence factor structure of the RIAS is based on Carroll's [29] three-stratum theory. On Stratum 1 the RIAS comprises two verbal subtests (Guess What, Verbal Reasoning) and two nonverbal subtests (Odd-Item Out, What's Missing?). On Stratum 2 the subtests are combined to create two factors, the Verbal Intelligence Index (VIX) and the Nonverbal Intelligence Index (NIX), which serve as indicators of crystallized and fluid intelligence, respectively. Stratum 3 comprises a Composite Intelligence Index (CIX), which is obtained by combining the VIX and NIX and which reflects general intelligence, *g*. A Composite Memory Index (CMX) on Stratum 2, which consists of two supplemental Stratum 1 subtests, is not integrated into the measure of general intelligence. On the basis of these theoretical assumptions, Reynolds and Kamphaus [22] suggested a two-factor intelligence structure with four subtests measuring two



**Fig 1. The Two-factor Structure (a) and Single-factor Structure (b) of the RIAS.** g = general intelligence; GWH = Guess What; VRZ = Verbal Reasoning; OIO = Odd-Item Out; WHM = What's Missing; e1-e4 = residuals; a = covariances of the residuals were set to be equal across groups.

doi:10.1371/journal.pone.0166533.g001

factors, verbal and nonverbal, (Fig 1a). Another structure assessed by Reynolds and Kamphaus [22] was a single-factor structure with four subtests measuring general intelligence (Fig 1b).

The RIAS intelligence factor structure proposed by Reynolds and Kamphaus [22] (Fig 1a) has been supported in the English, Danish, German, and Spanish RIAS versions by using confirmatory factor analysis (CFA) and standardization samples [22–25] as well as referred students in the English version [30]. Further, studies using CFA procedures also found support for the single-factor structure in the standardization samples of the four RIAS versions [31] as well as in typically developing and referred samples in the English version [32]. Additionally, studies using exploratory factor analysis have found evidence for a single-factor structure (Fig 1b) for the English RIAS standardization sample [33] and for referred samples [34–36]. While the theoretically proposed RIAS intelligence factor structure was supported in standardization and referred samples, it is unknown whether there is measurement invariance across groups with and without a migration background.

The main goal of the present study was to assess measurement invariance and mean differences in the German version of the RIAS in individuals with and without a migration background. The German version of the RIAS was standardized in Switzerland and Germany. Both countries have a high percentage of individuals with a migration background of over 20% of the population [2, 3]. Therefore, it is essential to ensure measurement invariance of diagnostic instruments such as the RIAS, so that examiners can rely on the validity of test results for individuals with a migration background. In the current study we addressed the following research questions: (a) Are the intelligence factor structures (i.e., single-factor structure and two-factor structure) of the RIAS invariant across individuals with and without a migration background, and (b) do individuals with and without a migration background differ in latent means and variances in the RIAS intelligence factors (the VIX, NIX, and CIX)?

First, we conducted single-group CFAs to determine model fit for the RIAS intelligence factor structures depicted in Fig 1a and 1b for both groups separately. Second, we ran a multi-

group CFA for both RIAS intelligence factor structures to assess measurement invariance across groups. We hypothesized that measurement invariance would hold across groups. When invariance held, we examined invariance of the latent variances and latent means across groups. We assumed that individuals with a migration background would have lower means in the latent variables VIX, NIX, and CIX than individuals without a migration background. We further assumed that lower latent means would in particular be evident in the RIAS VIX, drawing on the assumption that language barriers might have led to lower test performance as the verbal subtests require higher skills in German language than the nonverbal subtests [6–9].

## Method

### Participants

The study included 632 individuals with and without a migration background (each  $n = 316$ ) from the standardization sample of the German version of the RIAS (standardization sample  $N = 2,145$ ) [24]. This study was carried out in accordance with the recommendations of the Declaration of Helsinki, and it was approved by the Ethics Committee of Basel. Written informed consent was obtained from all subjects. For the children, parents gave written informed consent and assent was obtained from each child prior to the start of the study. Individuals with and without a migration background were matched in sex and age. Sex was equally distributed, with 48% female and 52% male, according to chi-square tests ( $p > .05$ ). Subjects were between the ages of 3 and 99 years ( $M = 15.79$ ,  $SD = 16.81$ ). The frequencies of individuals in different age ranges are provided in [S1 Table](#). Migration background was defined as speaking German as a first or a second language: Individuals without a migration background all spoke German as their first language, whereas individuals with a migration background all spoke German as a second language. In our study, 88% of individuals with a migration background spoke a language from a Western, European country as their first language, while 12% of the immigrants spoke a language from a non-Western country as their first language. First languages most often named by individuals with a migration background were Turkish (21%), Italian (8%), Serbian (6%), and Spanish (5%). [S2 Table](#) gives a detailed overview of all foreign first languages. The distribution of Western and non-Western immigrants represents the immigrant population in Switzerland and Germany well [2, 3], and thus all individuals with a migration background were included in the following analyses. However, there are studies that revealed group differences in intelligence across individuals from Western and immigrants from non-Western countries [37]. In the present study, the sample of immigrants from non-Western countries was not large enough to analyze such group differences. However, we also performed all analyses with immigrants from Western European countries only. As the results remained stable, we present results only for the whole sample including immigrants from both Western and non-Western countries, as we believe this distribution is more representative of the immigrant populations in Switzerland and Germany.

### Measures

The RIAS [22] is an individually administered intelligence test. Its German version [24] was standardized for individuals aged 3 to 99 years in Germany and Switzerland ( $N = 2,145$ ) in 2011 and 2012. The RIAS is composed of four intelligence subtests, which together constitute the CIX. The CIX can be further divided into two indices, represented by two subtests each: The verbal, VIX, comprises the subtests Guess What (GWH), in which individuals are asked to identify an object or a concept through the use of verbally presented clues, and Verbal Reasoning (VRZ), in which individuals are asked to complete verbal analogies. Therefore, the verbal subtests require high language skills. The nonverbal, NIX, comprises the subtests Odd-Item

Out (OIO), in which individuals are asked to identify a picture that does not go with others, and What's Missing? (WHM), in which individuals are required to identify the missing element in a presented picture. The CIX is based on the sum of the  $T$  scores ( $M = 50$ ;  $SD = 10$ ) of the four intelligence subtests. Reliability for the German RIAS is high for both groups with Cronbach's  $\alpha > .90$  for the subtests and  $\alpha > .95$  for intelligence indexes.

### Data analysis

To evaluate measurement invariance and invariance of latent variances and means across individuals with and without a migration background we used SPSS AMOS version 22 [38]. Correlation matrices with means and standard deviations of the subtest  $T$  scores were used as the input data file (see Table 1).

Following the suggestion of Meade, Johnson, and Braddy [39], we first conducted single-group CFAs to test the two-factor and single-factor structures, depicted in Fig 1, for each group. Second, we conducted a multi-group CFA for both factor structures across the two groups. Different levels of measurement invariance were tested, as described by Meredith [20], Milfont and Fischer [40], Steenkamp and Baumgartner [41], and Widaman and Reise [19]. First, we assessed configural invariance (Model 1). The intercepts of GWH and OIO were fixed to zero across both groups for the two-factor structure, and the intercept of GWH was

**Table 1. Correlations Between Reynolds Intellectual Assessment Scales Subtests and Indexes for Individuals With and Without a Migration Background With Means and Standard Deviations.**

Subtest or index	GWH <sup>a</sup>	VRZ <sup>a</sup>	OIO <sup>a</sup>	WHM <sup>a</sup>	VIX <sup>b</sup>	NIX <sup>b</sup>	CIX <sup>b</sup>
Without migration background							
GWH	1.00						
VRZ	.58	1.00					
OIO	.30	.35	1.00				
WHM	.32	.28	.38	1.00			
VIX	.88	.90	.36	.33	1.00		
NIX	.37	.38	.83	.83	.42	1.00	
CIX	.75	.77	.70	.68	.86	.83	1.00
Mean	53.11	53.39	52.21	51.60	105.85	103.50	105.34
SD	8.48	9.61	8.85	8.75	13.52	12.46	12.43
With migration background							
GWH	1.00						
VRZ	.65	1.00					
OIO	.37	.43	1.00				
WHM	.31	.32	.57	1.00			
VIX	.93	.88	.43	.34	1.00		
NIX	.38	.42	.89	.88	.44	1.00	
CIX	.77	.77	.78	.72	.85	.85	1.00
Mean	44.29	45.58	47.99	47.61	91.86	96.54	93.44
SD	11.28	8.96	10.66	10.28	15.39	15.68	14.92
Cohen's $d^c$	1.04	0.81	0.48	0.46	1.03	0.56	0.96

**Note.**  $N_{\text{Without migration background}} = 316$ ,  $N_{\text{With migration background}} = 316$ .  $SD$  = standard deviation; GWH = Guess What; VRZ = Verbal Reasoning; OIO = Odd-Item Out; WHM = What's Missing; VIX = Verbal Intelligence Index; NIX = Nonverbal Intelligence Index; CIX = Composite Intelligence Index.

<sup>a</sup> $T$  score normative mean = 50,  $SD = 10$ .

<sup>b</sup>Intelligence indexes normative mean = 100,  $SD = 15$ .

<sup>c</sup>Cohen's  $d$  is calculated based on the manifest data.

doi:10.1371/journal.pone.0166533.t001



fixed to zero across both groups for the single-factor structure. Second, we tested metric invariance (known as weak invariance; Model 2) by additionally constraining factor loadings across groups. Third, we examined scalar invariance (known as strong invariance; Model 3), by additionally constraining intercepts across groups. Fourth, we assessed residual invariance (known as strict invariance; Model 4) by additionally constraining error variances across groups. Finally, when scalar invariance held, invariance of the latent factors' variances and means could be analyzed [19] by constraining factor variances (Model 5) and factor means (Model 6) using the scalar invariance model as baseline.

For the single-factor structure, following Beaujean and McGlaughlin [34], we let the residuals of GWH and VRZ as well as the residuals of OIO and WHM to covary and constrained these two covariances to be equal [36]. We stress that this structure leads to a model empirically equivalent [42] to the two-factor structure as well as to a bifactor structure with a single general factor and two specific factors each measured by two indicators with factor loadings fixed to 1 [43, 44].

Following Hu and Bentler [45], we used a Comparative Fit Index (CFI) of  $\geq .95$ , a McDonald's Noncentrality Index (Mc) of  $\geq .90$ , and a root mean square error of approximation (RMSEA) of  $\leq .06$  to evaluate good model fit. Because chi-square test statistics tend to be sensitive to large sample sizes [39], we relied on the test of small difference in fit with  $p < .05$  [46] and changes in goodness-of-fit indices with  $\Delta\text{CFI} \leq .002$  and  $\Delta\text{Mc} \leq .005$  [39] to compare the nested models. We accepted model invariance if two of the three difference in fit statistics were within cutoff points.

Additionally, we computed Cohen's  $d$  [47] to investigate the effect sizes of the manifest and latent mean differences. A value of  $d = 0.2$  is considered a small effect,  $d = 0.5$  a medium effect, and  $d = 0.8$  a large effect.

## Results

### Single-group CFA

The two-factor (Fig 1a) and single-factor (Fig 1b) RIAS intelligence structures were evaluated for both groups using single-group CFA. As can be seen in Table 2, both structures yielded a good model fit for each group ( $\text{CFI} \geq .993$ ,  $\text{Mc} \geq .998$ ,  $\text{RMSEA} \leq .070$ ), indicating the models to be empirically equivalent [42]. Therefore, both structures were further analyzed to assess invariance of the latent variances and means of verbal and nonverbal intelligence (i.e., for this we analyzed the two-factor structure), as well as general intelligence (i.e., for this we analyzed

**Table 2. Fit Indices for Single-group Confirmatory Factor Analyses Evaluating Two Reynolds Intellectual Assessment Scales Intelligence Factor Structures in Individuals With and Without a Migration Background.**

Structure	$df$	$\chi^2$	CFI	Mc	RMSEA	90% CI
Two factors <sup>a</sup>						
Without migration background	1	2.556	.993	.998	.070	[.000, .183]
With migration background	1	0.438	1.000	1.000	< .001	[.000, .129]
Single factor <sup>a</sup>						
Without migration background	1	2.556	.993	.998	.070	[.000, .183]
With migration background	1	0.438	1.000	1.000	< .001	[.000, .129]

*Note.*  $N_{\text{without migration background}} = 316$ ,  $N_{\text{with migration background}} = 316$ . CFI = Comparative Fit Index; Mc = McDonald's Noncentrality Index; RMSEA = root mean square error of approximation; CI = confidence interval.

<sup>a</sup>Includes four subtests.

doi:10.1371/journal.pone.0166533.t002

the single-factor structure). The factor loadings, covariances, and correlations for both RIAS structures are provided in [S3 Table](#).

### Multi-group CFA

Measurement invariance across both groups was analyzed using multi-group CFA. For the two-factor structure ([Fig 1a](#)) results are shown in [Table 3](#). Configural, metric, and scalar invariance did hold across both groups with regard to  $\Delta CFI$ ,  $\Delta Mc$ , and the  $p$  value of the test of small difference in fit. Residual invariance yielded a poor fit regarding  $\Delta CFI$  and  $\Delta Mc$ , though the  $p$  value of the test of small difference in fit was  $< .05$ . Hence, we conclude that the variances of the residuals were noninvariant across groups.

Further, invariance of the latent variances and means for verbal and nonverbal intelligence (two-factor structure) were tested using the model implying scalar invariance as baseline (see [Table 3](#)). On the level of scalar invariance, results regarding verbal intelligence indicate that latent variances were invariant between groups with  $s^2 = 117.20$  for individuals with a migration background and  $s^2 = 72.00$  for individuals without a migration background. Latent means differed between groups and were lower for individuals with a migration background ( $M = 44.44$ ) compared to individuals without a migration background ( $M = 53.05$ ) with an effect size of  $d = 1.01$ , indicating a large effect. Regarding nonverbal intelligence, latent variances and latent means differed between groups. Latent variances were higher and latent means lower in individuals with a migration background ( $s^2 = 107.40$ ,  $M = 47.91$ ) compared to individuals without a migration background ( $s^2 = 72.81$ ,  $M = 52.30$ ). According to Cohen [47], the effect size of the group difference for nonverbal intelligence was medium with  $d = 0.52$ .

Results of the single-factor structure ([Fig 1b](#)) are presented in [Table 4](#). Though configural and metric invariance held, the models implying scalar and residual invariance showed values of  $\Delta CFI$  and  $\Delta Mc$  above the suggested cutoff values. Although the  $p$  value of the test of small

**Table 3. Fit Indices for Multi-group Confirmatory Factor Analysis Evaluating Measurement Invariance of the Two-factor Structure Including the Four Intelligence Subtests Across Individuals With and Without a Migration Background.**

Model	df	$\chi^2$	CFI	Mc	RMSEA	90% CI	$\Delta df$	$\Delta \chi^2$	$\Delta CFI$	$\Delta Mc$	Small diff $p$
1 Configural invariance	2	2.994	.998	.999	.028	[.000, .089]	—	—	—	—	—
2 Metric invariance	4	6.010	.997	.998	.028	[.000, .071]	—	—	—	—	—
2 versus 1	—	—	—	—	—	—	2	3.016	.001	< .001	.221
3 Scalar invariance	6	7.309	.998	.999	.019	[.000, .057]	—	—	—	—	—
3 versus 2	—	—	—	—	—	—	2	1.299	< .001	< .001	.522
4 Residual invariance	10	37.877	.954	.978	.067	[.045, .090]	—	—	—	—	—
4 versus 3	—	—	—	—	—	—	4	30.567	.044	.021	.196
5a Factor variances verbal <sup>a</sup>	7	12.273	.991	.996	.035	[.000, .066]	—	—	—	—	—
5a versus 3	—	—	—	—	—	—	1	4.963	.007	.003	.296
5b Factor variances nonverbal <sup>a</sup>	7	24.161	.971	.986	.062	[.036, .090]	—	—	—	—	—
5b versus 3	—	—	—	—	—	—	1	16.852	.027	.013	.217
6a Factor means verbal <sup>a</sup>	7	139.309	.780	.900	.173	[.149, .199]	—	—	—	—	—
6a versus 3	—	—	—	—	—	—	1	132.000	.218	.099	.031
6b Factor means nonverbal <sup>a</sup>	7	43.862	.939	.971	.091	[.067, .118]	—	—	—	—	—
6b versus 3	—	—	—	—	—	—	1	36.553	.059	.028	.145

*Note.*  $N_{\text{total}} = 632$ ,  $N_{\text{without migration background}} = 316$ ,  $N_{\text{with migration background}} = 316$ . CFI = Comparative Fit Index; Mc = McDonald's Noncentrality Index; RMSEA = root mean square error of approximation; CI = confidence interval.

<sup>a</sup>Factor variance invariance and factor mean invariance were calculated at the level of scalar invariance.

doi:10.1371/journal.pone.0166533.t003

**Table 4. Fit Indices for Multi-group Confirmatory Factor Analysis Evaluating Measurement Invariance of the Single-factor Structure Including the Four Intelligence Subtests Across Individuals With and Without a Migration Background.**

Model	df	$\chi^2$	CFI	Mc	RMSEA	90% CI	$\Delta df$	$\Delta \chi^2$	$\Delta CFI$	$\Delta Mc$	Small diff p
1 Configural invariance	3	11.506	.986	.993	.067	[.029, .110]	—	—	—	—	—
2 Metric invariance	6	16.481	.983	.992	.053	[.023, .084]	—	—	—	—	—
2 versus 1	—	—	—	—	—	—	3	4.975	.003	.001	.174
3 Scalar invariance	9	29.839	.965	.984	.061	[.037, .085]	—	—	—	—	—
3 versus 2	—	—	—	—	—	—	3	13.358	.018	.008	.271
4 Residual invariance	13	62.525	.918	.962	.078	[.059, .098]	—	—	—	—	—
4 versus 2	—	—	—	—	—	—	4	32.686	.047	.022	.181
5 Factor variances <sup>a</sup>	10	38.824	.952	.977	.068	[.046, .091]	—	—	—	—	—
5 versus 2	—	—	—	—	—	—	1	8.985	.013	.007	.270
6 Factor means <sup>a</sup>	10	150.616	.766	.895	.149	[.129, .171]	—	—	—	—	—
6 versus 2	—	—	—	—	—	—	1	120.777	.199	.089	.034

Note.  $N_{\text{total}} = 632$ ,  $N_{\text{without migration background}} = 316$ ,  $N_{\text{with migration background}} = 316$ . CFI = Comparative Fit Index; Mc = McDonald's Noncentrality Index; RMSEA = root mean square error of approximation; CI = confidence interval.

<sup>a</sup> Factor variance invariance and factor mean invariance were calculated at the level of measurement invariance, as scalar invariance did not hold.

doi:10.1371/journal.pone.0166533.t004

difference in fit is  $< .05$ , we conclude that there are differences in subtests' intercepts and residual variances across groups. As scalar invariance did not hold, we examined partial scalar invariance to identify specific subtests that caused noninvariance on the scalar level as proposed by Steenkamp and Baumgartner [41]. We found that the subtest OIO was the major contributor to the worsened model fit (S4 Table). Variances and means of general intelligence were not analyzed due to the lack of full scalar invariance [20].

## Discussion

The main goal of this study was to assess measurement invariance and differences in latent means of the German version of the RIAS in individuals with and without a migration background. We analyzed a two-factor structure with verbal and nonverbal intelligence as latent factors as well as a single-factor structure with general intelligence as the latent factor.

Both the two-factor and the single-factor structure were supported through single-CFAs for each group separately. This is in line with the RIAS technical manual [22, 24] and with previous research supporting the two-factor structure [30], the single-factor structure [23–26], or both RIAS intelligence factor structures [31, 32].

Then, we conducted multi-group CFAs to analyze measurement invariance for both structures across individuals with and without a migration background. For the two-factor structure, results showed scalar invariance across groups, indicating that differences in the subtest means were due to differences in the means of the underlying constructs [48]. Residual invariance could not be supported, possibly because of differing reliabilities of the scales across groups [49].

Further, we assessed invariance of latent variances and means of verbal and nonverbal intelligence. Results indicate that variances of nonverbal intelligence as well as means of verbal and nonverbal intelligence were noninvariant across groups. Individuals with a migration background showed similar variances in verbal intelligence and higher variances in nonverbal intelligence compared to individuals without a migration background. These results suggest that individuals with a migration background showed about the same range in verbal intelligence test scores but a wider range in nonverbal intelligence test scores than individuals

without a migration background. Beaujean and McGlaughlin [34] found a similar pattern when examining referred black and white students, but black students showed a higher variance in general intelligence in the RIAS test scores. Further, as hypothesized, individuals with a migration background achieved lower latent means in verbal as well as nonverbal intelligence when tested with the German version of the RIAS. This is in line with other studies that also found lower means in intelligence test scores for individuals with a migration background [5, 7, 9]. For the group differences, the effect size in verbal intelligence was high, whereas the effect size in nonverbal intelligence was medium, according to Cohen [47]. This is in accordance with our hypothesis that differences in latent means would be especially evident in the RIAS VIX. A possible explanation for this result may be that language barriers in individuals with a migration background might have led to lower test performance especially in the RIAS subtests with high language requirements [6, 8, 9]. However, individuals with a migration background also showed lower test performance in nonverbal intelligence. Existing literature suggests that not only language barriers but also cultural and environmental differences such as familiarity with tasks used in intelligence tests [11, 12], stereotype threat [13, 18], and parental language skills [14] can contribute to lower test performance in individuals with a migration background [10], which also might have led to the lower test scores of individuals with a migration background in the verbal and nonverbal intelligence indexes in the present study. In sum, the results of the present study indicate that using the German version of the RIAS may lead to underestimated intelligence test scores in individuals with a migration background. Such an approach in which examiners focus exclusively on the test results is called a “static” intelligence assessment ([12], p. 445). In contrast, “dynamic” testing methods ([12], p. 445) may constitute an assessment approach that is able to account for differences in individuals from different cultural contexts. In line with the zone of proximal development concept [50], in dynamic testing, the tasks are first introduced to familiarize the individuals with the nature of these tasks. Afterward the improvement in these tasks is assessed to evaluate the learning potential of an examinee, which may be a better estimate of abilities compared to results gained from a static assessment [51]. Thus, there is evidence that dynamic testing leads to diminishing test score differences across individuals with and without a migration background [11, 12]. These results indicate that examiners in psychological practice should consider using a dynamic testing approach to reduce effects of familiarity with intelligence test tasks. It should be a goal for future studies to determine whether the positive effects of such a dynamic approach are also evident when using the German version of the RIAS in individuals with and without a migration background.

For the single-factor structure, only metric invariance and not scalar invariance held. This indicates that the relationship between the observed scores and the latent construct was different across groups [20]. Additionally, we assessed partial scalar invariance and found that the major contributor to differences across groups was the subtest OIO. This is in line with other studies that also found OIO to be the major contributor to a lack of full scalar invariance [30] or a decreased model fit of scalar invariance [34]. Thus, group differences in the latent means can be explained either through the mean differences in the subtest OIO or through true latent mean differences [48].

Because full scalar invariance did not hold for the single-factor structure, we did not analyze invariance of latent variances and means of general intelligence [21, 41]. This leads to the conclusion that test scores of individuals with a migration background should not be compared with test scores of individuals without a migration background on the level of general intelligence when tested with the RIAS, but rather on the level of verbal and nonverbal intelligence only.

There are strengths and limitations of this study. We see it as a strength that we analyzed competing factor structures (two-factor vs. single-factor) of the RIAS, and thus our results contribute to the discussion on the best-fitting RIAS intelligence factor structure for specific samples. Further, we analyzed large age- and sex-matched samples of individuals with and without a migration background. However, we could not control for social factors such as socioeconomic status (SES); previous studies have shown that SES is positively correlated with intelligence [52] and negatively related to migration background such that individuals with a migration background have a lower SES than individuals without a migration background [53, 54]. Therefore, it is possible that the means in individuals with a migration background would have been slightly higher if we additionally could have controlled for SES. Further, we determined migration background solely on the basis of individuals being foreign-language or native. Future studies could additionally gather the status of migration (country of birth, number of the family's generations living in the country of immigration, etc.). Finally, our sample of individuals with a migration background included subjects with a diversity of first languages. Although this reflects the language diversity of Switzerland and Germany [2, 3], it can be assumed that these individuals do not represent a homogeneous group. Future studies might therefore focus on specific groups such as individuals from countries with collectivist or individualist cultural values [55] or from different language families [56]. Future studies might also analyze RIAS test scores in individuals with and without a migration background in other countries with a high percentage of migration.

## Conclusion

In conclusion, the RIAS two-factor intelligence structure including four subtests measuring verbal and nonverbal intelligence showed measurement invariance across samples, whereas we could only demonstrate partial scalar invariance for the single-factor structure with four subtests measuring general intelligence. These results suggest that examiners should focus on the VIX and NIX and not on the CIX when comparing RIAS test results of individuals with and without a migration background. Finally, examiners should consider the migration background as well as cultural, environmental, and social factors of an examinee when interpreting RIAS test results, as individuals with a migration background may achieve lower test scores in particular on the VIX but also on the NIX than individuals without a migration background, which may be due to language, cultural, environmental, or social barriers.

## Supporting Information

### S1 Table. Frequencies for Age Ranges for the Total Sample.

(DOCX)

### S2 Table. Overview of All First Languages Named by Subjects With a Migration Background.

(DOCX)

### S3 Table. Unstandardized and Standardized Factor Loadings, Covariances, and Correlations for the Two-Factor and Single-Factor Structures of the Confirmatory Factor Analyses for Individuals With and Without a Migration Background.

(DOCX)

### S4 Table. Fit Indices for Multi-group Confirmatory Factor Analysis Evaluating Partial Measurement Invariance of the Single-factor Structure Including the Four Intelligence Subtests Across Individuals With and Without a Migration Background.

(DOCX)

## Author Contributions

**Conceptualization:** JTG EF AG PH.

**Data curation:** JTG EF PH.

**Formal analysis:** JTG EF PH.

**Funding acquisition:** AG PH.

**Investigation:** JTG EF AG PH.

**Methodology:** JTG EF PH AG.

**Project administration:** AG PH.

**Resources:** AG PH.

**Supervision:** PH.

**Visualization:** JTG.

**Writing – original draft:** JTG PH.

**Writing – review & editing:** JTG EF AG PH.

## References

1. Eurostat. The EU in the world—population; 2014. Available: [http://ec.europa.eu/eurostat/statistics-explained/index.php/The\\_EU\\_in\\_the\\_world\\_-\\_population#Migration\\_and\\_asylum](http://ec.europa.eu/eurostat/statistics-explained/index.php/The_EU_in_the_world_-_population#Migration_and_asylum)
2. Bundesamt für Statistik. Migration und Integration; 2015. Available: <http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/07/blank/key/01/01.htmlx>
3. Statistisches Bundesamt. Bevölkerungsstand; 2015. Available: <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Bevoelkerungsstand/Bevoelkerungsstand.html;jsessionid=D9339CCA1586CB4B47CA4D661CFD1EE8.cae1>
4. Goldstein S, Princiotta D, Naglieri JA, editors. Handbook of intelligence. Evolutionary theory, historical perspective, and current concepts. New York, NY: Springer; 2015. doi: [10.1007/978-1-4939-1562-0](https://doi.org/10.1007/978-1-4939-1562-0)
5. Calero MD, Fernández-Parra A, López-Rubio S, Carles R, Mata S, Vives C, et al. Variables involved in personal, social and school adjustment in a sample of preschool-aged children from different cultural backgrounds. Eur J Psychol Educ. 2013; 28:133–155. doi: [10.1007/s10212-012-0107-8](https://doi.org/10.1007/s10212-012-0107-8)
6. Daseking M, Lipsius M, Petermann D, Waldmann H-C. Application of the German WISC-IV in children with diverse cultural and linguistic backgrounds. Kindheit und Entwicklung. 2008; 17:76–89.
7. Hagmann-von Arx P, Petermann F, Grob A. Konvergente und diskriminante Validität der WISC-IV und der Intelligence and Development Scales (IDS) bei Kindern mit Migrationshintergrund [Convergent and discriminant validity of the WISC-IV and the Intelligence and Development Scales (IDS) in children with a migration background]. Diagnostica. 2013; 59:170–182. doi: [10.1026/0012-1924/a000091](https://doi.org/10.1026/0012-1924/a000091)
8. Weiss LG., Harris JG, Prifitera A, Courville T, Rolfhus E, Saklofske DH, et al. WISC-IV interpretation in societal context. In: Weiss LG, Saklofske DH, Prifitera A, Holdnack JA, editors. WISC-IV advanced clinical interpretation. San Diego, CA: Elsevier; 2006. pp. 1–57.
9. Harris JG, Muñoz MR, Llorente AM. Considerations in using the WISC-IV with Hispanic children. In: Prifitera A, Saklofske DH, Weiss LG, editors. WISC-IV Clinical assessment and intervention. San Diego, CA: Elsevier; 2008. pp. 409–442.
10. Mackintosh NJ. IQ and human intelligence. 2nd ed. New York: Oxford University Press; 2011.
11. Calero MD, Mata S, Carles R, Vives C, López-Rubio S, Fernández-Parra A, et al. Learning potential assessment and adaptation to the educational context: The usefulness of the ACFS for assessing immigrant preschool children. Psychol Sch. 2013; 50:705–721. doi: [10.1002/pits.21701](https://doi.org/10.1002/pits.21701)
12. Resing WC, Tunteler E, de Jong FM, Bosma T. Dynamic testing in indigenous and ethnic minority children. Learn Individ Differ. 2009; 19:445–450. doi: [10.1016/j.lindif.2009.03.006](https://doi.org/10.1016/j.lindif.2009.03.006)
13. Appel M, Weber S, Kronberger N. The influence of stereotype threat on immigrants: review and meta-analysis. Frontiers Psychol. 2015; 6:900. doi: [10.3389/fpsyg.2015.00900](https://doi.org/10.3389/fpsyg.2015.00900) PMID: [26217256](https://pubmed.ncbi.nlm.nih.gov/26217256/)

14. Lahaie C. School readiness of children of immigrants: Does parental involvement play a role? *Soc Sci Q*. 2008; 89:684–705.
15. Turney K, Kao G. Barriers to school involvement: Are immigrant parents disadvantaged? *J Educ Res*. 2009; 102:257–271. doi: [10.3200/JOER.102.4.257-271](https://doi.org/10.3200/JOER.102.4.257-271)
16. Hessels MG. Low IQ but high learning potential: Why Zeineb and Moussa do not belong in special education. *Educ Child Psychol*. 1997; 14:121–136.
17. Klingner JK, Blanchett WJ, Harry B. Race, culture, and developmental disabilities. In: Odom SL, Horner RH, Snell ME, editors. *Handbook of developmental disabilities*. New York: Guilford Press; 2007. pp. 55–75.
18. Baltes BB, Rudolph CW. Examining the effect of negative Turkish stereotypes on evaluative workplace outcomes in Germany. *J Manag Psychol*. 2010; 25:148–158. doi: [10.1108/02683941011019357](https://doi.org/10.1108/02683941011019357)
19. Widaman KF, Reise SP. Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In Bryant K, Windle M, West S, editors. *The science of prevention: Methodological advances from alcohol and substance abuse research*. Washington, DC: American Psychological Association; 1997. pp. 281–324. doi: [10.1037/10222-009](https://doi.org/10.1037/10222-009)
20. Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*. 1993; 48:525–543. doi: [10.1007/BF02294825](https://doi.org/10.1007/BF02294825)
21. Chen FF, Sousa KH, West SG. Testing measurement invariance of second-order factor models. *Struct Equ Modeling*. 2005; 12:471–492.
22. Reynolds CR, Kamphaus RW. *Reynolds Intellectual Assessment Scales: Professional manual*. Lutz, FL: PAR; 2003.
23. Hartmann P, Andresen J. Dokumentation af den danske standardisering af RIAS. Virum, Denmark: Hogrefe Psykologisk; 2011.
24. Hagmann-von Arx P, Grob A. *Reynolds Intellectual Assessment Scales (RIAS)*. Bern, Switzerland: Hans Huber; 2014.
25. Santamaría P, Fernández Pinto I. *RIAS Escalas de Inteligencia de Reynolds y RIST Test de Inteligencia Breve de Reynolds Manual. Adaptación española*. Madrid, Spain: TEA Ediciones; 2009.
26. Andrews JJW. *RIAS: Reynolds Intellectual Assessment Scales*. *J Psychoeduc Assess*. 2007; 25:402–408.
27. Dombrowski SC, Mrazik M. Test review: Reynolds Intellectual Assessment Scales. *Can J Sch Psychol*. 2008; 23:223–230.
28. Elliot RW. Test review: Reynolds Intellectual Assessment Scales. *Arch Clin Neuropsychol*. 2004; 19:325–328. PMID: [15046017](https://pubmed.ncbi.nlm.nih.gov/15046017/)
29. Carroll JB. *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press; 1993.
30. Beaujean AA, McGlaughlin SM, Margulies AS. Factorial validity of the Reynolds Intellectual Assessment Scales for referred students. *Psychol Sch*. 2009; 46:932–950. doi: [10.1002/pits.20435](https://doi.org/10.1002/pits.20435)
31. Gygi JT, Ledermann T, Grob A, Hagmann-von Arx P. The Reynolds Intellectual Assessment Scales (RIAS): Measurement invariance across four language groups. Manuscript submitted for publication; 2016.
32. Irwin JK, Joschko M, Kerns KA. Confirmatory factor analysis of the Reynolds Intellectual Assessment Scales (RIAS) in Canadian children. *Clin Neuropsychol*. 2014; 28:1258–1277. doi: [10.1080/13854046.2014.975843](https://doi.org/10.1080/13854046.2014.975843) PMID: [25359205](https://pubmed.ncbi.nlm.nih.gov/25359205/)
33. Dombrowski SC, Watkins MW, Brogan MJ. An exploratory investigation of the factor structure of the Reynolds Intellectual Assessment Scales (RIAS). *J Psychoeduc Assess*. 2009; 27:494–507. doi: [10.1177/0734282909333179](https://doi.org/10.1177/0734282909333179)
34. Beaujean AA, McGlaughlin SM. Invariance in the Reynolds Intellectual Assessment Scales for black and white referred students. *Psychol Assess*. 2014; 26:1394–1399. doi: [10.1037/pas0000029](https://doi.org/10.1037/pas0000029) PMID: [25285717](https://pubmed.ncbi.nlm.nih.gov/25285717/)
35. Nelson JM, Canivez GL. Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychol Assess*. 2012; 24:129–140. doi: [10.1037/a0024878](https://doi.org/10.1037/a0024878) PMID: [21823799](https://pubmed.ncbi.nlm.nih.gov/21823799/)
36. Nelson JM, Canivez GL, Lindstrom W, Hatt CV. Higher-order exploratory factor analysis of the Reynolds Intellectual Assessment Scales with a referred sample. *J Sch Psychol*. 2007; 45:439–456. doi: [10.1016/j.jsp.2007.03.003](https://doi.org/10.1016/j.jsp.2007.03.003)
37. te Nijenhuis J, Willigers D, Dragt J, van der Flier H. The effects of language bias and cultural bias estimated using the method of correlated vectors on a large database of IQ comparisons between native Dutch and ethnic minority immigrants from non-Western countries. 2016; 54:117–135.

38. Arbuckle JL. IBM SPSS Amos 22 user's guide. Armonk, NY: IBM; 2013.
39. Meade AW, Johnson EC, Braddy PW. Power and sensitivity of alternative fit indices in tests of measurement invariance. *J Appl Psychol*. 2008; 93:568–592. doi: [10.1037/0021-9010.93.3.568](https://doi.org/10.1037/0021-9010.93.3.568) PMID: [18457487](https://pubmed.ncbi.nlm.nih.gov/18457487/)
40. Milfont TL, Fischer R. Testing measurement invariance across groups: Applications in cross-cultural research. *Int J Psychol Res*. 2010; 3:111–130.
41. Steenkamp JBE, Baumgartner H. Assessing measurement invariance in cross-national consumer research. *J Consum Res*. 1998; 25:78–107. doi: [10.1086/209528](https://doi.org/10.1086/209528)
42. Lee S, Hershberger S. A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behav Res*. 1990; 25:313–334. doi: [10.1207/s15327906mbr2503\\_4](https://doi.org/10.1207/s15327906mbr2503_4) PMID: [26761406](https://pubmed.ncbi.nlm.nih.gov/26761406/)
43. Canivez GL, Watkins MW. Review of the Wechsler Intelligence Scale for Children—Fifth Edition: Critique, commentary, and independent analyses. In: Kaufman AS, Raiford SE, Coalson DL, editors. *Intelligent testing with the WISC–V*. Hoboken, NJ: Wiley; 2016. pp. 683–702.
44. Reise SP. The rediscovery of bifactor measurement models. *Multivariate Behav Res*. 2012; 47:667–696. doi: [10.1080/00273171.2012.715555](https://doi.org/10.1080/00273171.2012.715555) PMID: [24049214](https://pubmed.ncbi.nlm.nih.gov/24049214/)
45. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999; 6:1–55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
46. MacCallum RC, Browne MW, Cai L. Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychol Methods*. 2006; 11:19–35. doi: [10.1037/1082-989X.11.1.19](https://doi.org/10.1037/1082-989X.11.1.19) PMID: [16594765](https://pubmed.ncbi.nlm.nih.gov/16594765/)
47. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum; 1988.
48. Meredith W, Teresi JA. An essay on measurement and factorial invariance. *Medical Care*. 2006; 44:69–77. doi: [10.1097/01.mlr.0000245438.73837.89](https://doi.org/10.1097/01.mlr.0000245438.73837.89) PMID: [17060838](https://pubmed.ncbi.nlm.nih.gov/17060838/)
49. DeShon RP. Measures are not invariant across groups without error variance homogeneity. *Psychol Sci*. 2004; 46:137–149.
50. Vygotsky LS. Interaction between learning and development (M. Lopez-Morillas, Trans.). In: Cole M, John-Steiner V, Scribner S, Souberman E, editors. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press; 1978. pp. 79–91.
51. Sternberg RJ, Grigorenko EL. *Dynamic testing: The nature and measurement of learning potential*. New York: Cambridge University Press; 2002.
52. Bradley RH, Corwyn RF. Socioeconomic status and child development. *Annu Rev Psychol*. 2002; 53:371–399. doi: [10.1146/annurev.psych.53.100901.135233](https://doi.org/10.1146/annurev.psych.53.100901.135233) PMID: [11752490](https://pubmed.ncbi.nlm.nih.gov/11752490/)
53. Bundesamt für Statistik. Verteilung des verfügbaren Äquivalenzeinkommens und das Quintilverhältnis S80/S20, nach verschiedenen soziodemografischen Merkmalen; 2014. Available: <http://www.bfs.admin.ch/bfs/portal/de/index/themen/20/03/blank/key/02/06.html>.
54. DeNavas-Walt C, Proctor BD. Income and poverty in the United States: 2014. U.S. Census Bureau Current Population Reports, P60-252; 2015. Available: <http://www.census.gov/content/dam/Census/library/publications/2015/demo/p60-252.pdf>
55. Hofstede G. *Culture's consequences comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage; 2001.
56. Ruhlén M. *A guide to the world's languages: Volume I, Classification*. Stanford, CA: Stanford University Press; 1991.



**APPENDIX C: Study 3**

Gygi, J. T., Hagmann-von Arx, P., Schweizer, F., & Grob, A. The predictive validity of four intelligence tests for school grades: Results of a longitudinal study in a small sample of typically developing children. *Manuscript submitted for publication to Frontiers in Psychology*.

Draft December 1, 2016

Running Head: INTELLIGENCE AND SCHOOL GRADES

**The predictive validity of four intelligence tests for school grades:  
Results of a longitudinal study in a small sample of typically  
developing children**

Jasmin T. Gygi\*, Priska Hagmann-von Arx, Florine Schweizer, and Alexander Grob  
Department of Psychology, University of Basel, Basel, Switzerland

\*Correspondence:  
Jasmin T. Gygi  
jasmin.gygi@unibas.ch

## INTELLIGENCE AND SCHOOL GRADES

## Abstract

Intelligence is considered the strongest single predictor of scholastic achievement. However, little is known regarding the predictive validity of well-established intelligence tests for school grades. We analyzed the predictive validity of four widely used intelligence tests in German-speaking countries: The Intelligence and Development Scales (IDS), the Reynolds Intellectual Assessment Scales (RIAS), the Snijders-Oomen Nonverbal Intelligence Test (SON-R 6-40), and the Wechsler Intelligence Scale for Children (WISC-IV), which were individually administered to 103 typically developing children ( $M_{\text{age}} = 9.17$  years) enrolled in regular school. School grades were collected longitudinally after 3 years (averaged school grades, mathematics, and language) and were available for 54 children ( $M_{\text{age}} = 11.77$  years). All four tests significantly predicted averaged school grades ( $\beta = .259$  to  $.366$ ,  $p < .01$  to  $.05$ ). Furthermore, the IDS predicted both mathematics ( $\beta = .341$ ,  $p < .05$ ) and language ( $\beta = .307$ ,  $p < .05$ ). The SON-R 6-40 predicted mathematics ( $\beta = .273$ ,  $p < .05$ ), while the RIAS predicted language ( $\beta = .322$ ,  $p < .01$ ). The WISC-IV showed no significant association with longitudinal scholastic achievement when mathematics and language were analyzed separately. The results revealed the predictive validity of currently used intelligence tests for longitudinal scholastic achievement in German-speaking countries and support their use in psychological practice, in particular for predicting averaged school grades. However, the results have to be considered as preliminary because only a small sample of typically developing children enrolled in regular school could be analyzed.

*Keywords:* validity, scholastic achievement, IDS, RIAS, SON-R 6-40, WISC-IV

## INTELLIGENCE AND SCHOOL GRADES

**Introduction**

The primary purpose of the first intelligence test (Binet and Simon, 1905) was to predict scholastic achievement in order to determine the best school setting for a child. Since the beginning of intelligence assessment, the predictive validity of intelligence test scores for scholastic achievement has been well studied. Cross-sectional and longitudinal studies indicated strong correlations, around  $r = .40$  to  $.81$ , between the two (e.g., Deary, Strand, Smith, and Fernandes, 2007; Mackintosh, 2011; Sternberg, Grigorenko, and Bundy, 2001). The association between intelligence and scholastic achievement seems to be stronger when using standardized achievement tests compared to school grades (Rost, 2009; Sternberg et al., 2001). Standardized achievement tests represent achievement at only one point in time, whereas school grades represent achievement over a longer period and thus may also be influenced by other constructs such as self-control and motivation (Rost, 2009). However, school grades are crucial for children to be promoted to the next higher grade level as well as for further scholastic and occupational qualifications (Roth, Becker, Romeyke, Schäfer, Domnick, and Spinath, 2015).

Focusing on school grades, a recent meta-analysis (Roth et al., 2015) found an observed correlation of  $r = .44$  and an estimated true correlation (i.e., corrected for error of measurement and range restriction) of  $\rho = .54$  between intelligence and school grades. Regarding subject domains, the correlations were highest and comparable for mathematics/science ( $r = .42$ ,  $\rho = .49$ ) and languages ( $r = .36$ ,  $\rho = .44$ ). The results furthermore revealed that correlations between intelligence and school grades in elementary school ( $r = .40$ ,  $\rho = .45$ ) tended to be weaker than in middle and high school ( $r = .46$ ,  $\rho = .54$  to  $.58$ ), because intelligence deficits in elementary school may be compensated more easily through practice than in higher-grade levels, as the learning content is easier to understand. This result is in contrast to previous research (e.g., Sternberg et al., 2001), that identified stronger correlations between intelligence and scholastic achievement in elementary school than in higher-grade levels, because of growing range restrictions.

The meta-analysis performed by Roth et al. (2015) included studies conducted in different countries. In German-speaking countries, for example, the Culture Fair Test-20-Revision (Weiss, 2006), standardized in 2003, showed associations with school grades in mathematics/science ranging from  $r = .26$  to  $.39$  and in languages of  $r = .23$ . Further, the German Cognitive Ability Test – 4-12 – Revision (Heller and Perleth, 2000), standardized from 1995 to 1997, showed associations with school grades in mathematics/science ranging from  $r = .17$  to  $.60$  and in languages ranging from  $r = .12$  to  $.14$ . However, the meta-analysis did not include more recently standardized intelligence tests currently used in German-speaking countries.

Currently used intelligence tests in German-speaking countries (e.g., Hagmann-von Arx, Gauck, and Grob, 2015) include (a) the Intelligence and Development Scales (IDS; Grob, Meyer, and Hagmann-von Arx, 2013), an intelligence test for children aged 5 to 10 years measuring in particular fluid intelligence; (b) the Reynolds Intellectual Assessment Scales (RIAS; Reynolds and Kamphaus, 2003; German version: Hagmann-von Arx and Grob, 2014), an intelligence test for individuals aged 3 to above 90 years that measures verbal and nonverbal intelligence, based on crystallized and fluid intelligence, respectively. A composite intelligence index can be computed from the values in verbal and nonverbal intelligence; (c) the Snijders-Oomen Nonverbal Intelligence Test Revised 6-40 (SON-R 6-40; Tellegen, Laros, and Petermann, 2012), a nonverbal intelligence test measuring fluid intelligence in individuals aged 6 to 40 years; and (d) the Wechsler Intelligence Scales for Children, Fourth Edition (WISC-IV; Wechsler, 2003; German version: Petermann and Petermann, 2011), an intelligence test used worldwide to measure general intelligence (Full-Scale IQ or FSIQ). Additionally, the WISC-IV provides four index scores: verbal comprehension reflecting the understanding of verbal concepts; perceptual reasoning measuring nonverbal perception and

## INTELLIGENCE AND SCHOOL GRADES

manipulation; working memory assessing attention and working memory; and processing speed reflecting visuospatial speed of processing.

Nevertheless, little is known regarding the predictive validity of these intelligence tests for school grades. Especially for German-speaking countries and for studies independent of the standardization samples, there is a lack of literature analyzing predictive validity of these tests. For the IDS, Gut, Reimann, and Grob (2013) analyzed the predictive validity of general intelligence in children aged 5 to 7 years from the standardization sample for concurrent ( $n = 402$ ) and longitudinal ( $n = 221$ ) scholastic achievement. Concurrent scholastic achievement was operationalized through parents' and teachers' ratings in mathematics, science, and language (German), which were averaged across subjects. Longitudinal scholastic achievement was based on averaged school grades in these subjects 3 years later. Results revealed medium to large effect sizes ( $\beta = .30$  to  $.56$ ) for the cross-sectional and a small effect size ( $\beta = .21$ ) for the longitudinal association. These results replicate findings of a prior study conducted by Gut et al. (2012) showing that in an extended sample of 263 children aged 5 to 10 years, IDS general intelligence predicted school grades in mathematics, science, and language (German) 3 years later with medium effect sizes ( $\beta = .28$  to  $.34$ ). Both studies indicate small to moderate concurrent and predictive validity of the IDS general intelligence for averaged school grades.

For the German version of the RIAS, we found no studies on the predictive validity of intelligence indices on school grades. However, the Technical Manual of the English Version of the RIAS (Reynolds and Kamphaus, 2003) reports a cross-sectional validation study conducted with 78 children aged 3 to 16 years. Results revealed strong correlations between the composite intelligence index and a standardized achievement test in mathematics ( $r = .67$ ) and language ( $r = .64$ ), indicating strong concurrent validity of the RIAS composite intelligence index for standardized achievement tests.

For the SON-R 6-40, the Technical Manual of the German version (Tellegen et al., 2012) reports moderate to strong correlations between the test scores and concurrent school grades in mathematics ( $r = .58$ ) and language ( $r = .49$ ) for 182 elementary school children aged 6 to 11 years. These results indicate that nonverbal intelligence measured using the SON-R 6-40 shows moderate to strong concurrent validity for school grades.

For the German version of the WISC-IV there are, to our knowledge, no available studies on the predictive validity for school grades. For the English version of the WISC-IV, Glutting, Watkins, Konold, and McDermott (2006) studied the concurrent validity of general intelligence and its specific indices on a standardized academic achievement test in mathematics and reading with a sample of 498 individuals aged 6 to 16 years from the WISC-IV standardization sample. Results showed large effect sizes for the WISC-IV FSIQ (60% variance explained) but only small effect sizes for the specific indices (0 to 2% additional variance explained). The FSIQ predicted concurrent mathematics and reading equally well. Thus, results indicate that in particular the WISC-IV FSIQ is correlated with concurrent standardized academic achievement tests.

In sum, Roth and colleagues' (2015) meta-analysis revealed that intelligence test scores correlate moderately to strongly with school grades in mathematics and language. However, there is only scarce evidence regarding the longitudinal prediction of school grades with currently used intelligence tests in German-speaking countries.

In the current study we analyzed the predictive power of the German versions of the IDS, the RIAS, the SON-R 6-40, and the WISC-IV for longitudinal school grades. We analyzed the general intelligence indices only, as the Technical Manuals (e.g., Reynolds and Kamphaus, 2003) as well as previous research (Glutting et al., 2006) do not recommend the use of specific indices for high-stakes decisions because of lowered reliability and validity compared to the general intelligence indices. On the basis of previous research (e.g., Deary et al., 2007; Glutting et al., 2006; Gut et al., 2012, 2013, Roth et al., 2015), we expected that the

## INTELLIGENCE AND SCHOOL GRADES

general intelligence indices would positively predict averaged school grades as well as school grades in mathematics and language (German) with medium to strong effect sizes.

### Materials and Methods

#### Participants

The sample consisted of 103 typically developing children aged 6 to 11 years ( $M = 9.17$  years,  $SD = 0.93$ ; 52% females, 48% males) enrolled in regular schools. All children took part in an intelligence assessment. Three years later, parents of 54 children aged 10 to 13 years ( $M = 11.77$  years,  $SD = 0.79$ ; 52% females, 48% males) provided information about their children's school grades in mathematics and language. Regarding parental education, 74% of the parents had a non-tertiary education and 26% had a tertiary education. This distribution indicates that parent's educational attainment in the present study is comparable with the general Swiss population (Swiss Federal Statistical Office, 2016). Post hoc power analysis using G\*Power (Faul, Erdfelder, Lang, and Buchner, 2007) indicated that with a chance of 80% and a .05 alpha level, the current study was sufficiently powered to detect medium effect sizes ( $r = .30$ ; Cohen, 1988). The 54 children who participated in both study waves showed significantly higher intelligence scores in the RIAS composite intelligence index ( $M = 103.24$ ,  $SD = 8.29$ ) as well as in the WISC-IV FSIQ ( $M = 107.39$ ,  $SD = 10.58$ ) than the 49 children who did not participate in the second study wave (RIAS:  $M = 98.98$ ,  $SD = 9.19$ ,  $F = 0.82$ ,  $p < .01$ ; WISC-IV:  $M = 102.12$ ,  $SD = 12.78$ ,  $F = 1.57$ ,  $p < .05$ ). No differences were found for the IDS and the SON-R 6-40.

#### Measures

**Intelligence.** The IDS assesses general intelligence and five developmental domains (psychomotor skills, social-emotional competences, mathematics, language, and achievement motivation) in children aged 5 to 10 years. For the current study, only general intelligence was analyzed. IDS general intelligence consists of seven subtests (i.e., visual perception, selective attention, phonological memory, visual-spatial memory, auditory memory, abstract reasoning, figural reasoning), which measure mainly fluid intelligence. The administration of IDS general intelligence takes about 45 minutes. The IDS was standardized from 2007 to 2008 in Austria, Germany, and Switzerland. Reliability for general intelligence is high with Cronbach's  $\alpha = .92$ .

The RIAS is an intelligence test for individuals aged 3 to above 90 years. It comprises four intelligence subtests (i.e., guess what, verbal reasoning, odd-item out, what's missing), which together constitute the composite intelligence index, CIX. The CIX can also be divided into two indices, represented by two of the four above mentioned subtests each: the Verbal Intelligence Index, VIX, representing crystallized intelligence, and the Nonverbal Intelligence Index, NIX, representing fluid intelligence. Two additional subtests can be administered measuring verbal and nonverbal memory resulting in a Composite Memory Index. The memory subtests are not entered in the CIX. The assessment of the RIAS CIX takes about 20 to 25 minutes. The German version of the RIAS was standardized from 2011 to 2012 in Germany and Switzerland. Reliability for the RIAS is high with Cronbach's  $\alpha = .95$  for the CIX and  $\alpha = .93$  to  $.94$  for the VIX and NIX.

The SON-R 6-40 assesses nonverbal intelligence for individuals aged 6-40 years. It comprises four subtests (i.e., analogies, categories, mosaics, patterns) that primarily measure fluid intelligence. The administration of the SON-R 6-40 takes about 45 to 60 minutes. The German version of the SON-R 6-40 was standardized from 2009 to 2011 in Germany and the Netherlands. Reliability for the SON-R 6-40 is high with Cronbach's  $\alpha = .95$ .

The WISC-IV is an intelligence test measuring general intelligence for children aged 6 to 16 years. It includes 10 core subtests (i.e., similarities, vocabulary, comprehension, block design, picture concepts, matrix reasoning, digit span, letter-number sequencing, symbol

## INTELLIGENCE AND SCHOOL GRADES

search, coding) that constitute the FSIQ and four specific indices: the Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index. The administration of the WISC-IV core subtests takes about 60 minutes. The German version of the WISC-IV was standardized from 2005 to 2006 in Austria, Germany, and Switzerland. Reliability for the WISC-IV is high with  $r = .97$  for the FSIQ and  $r = .87$  to  $.94$  for the specific intelligence indices.

**School grades.** Three years after intelligence assessment, parents were asked to report on their child's school grades in mathematics and language (1 = poorest grade, 6 = best grade; grades 4 to 6 represent the passing range) based on the school records of the latest term (i.e., overall grades). In Switzerland, passing grades in both mathematics and language are crucial for a child to be promoted to the next higher grade level (Swiss Media Institute for Education and Culture, 2016). Thus, in line with previous research (e.g., Gut et al., 2013) school grades were additionally averaged across subjects to obtain a composite estimate of scholastic achievement.

### Procedure

This study was carried out in accordance with the recommendations of the Ethics Committee of Basel, Switzerland and with the Declaration of Helsinki. Parents gave written informed consent prior to participation in the study, and assent was obtained from the children. Children were recruited from elementary schools in the German-speaking part of Switzerland in 2011. Trained study personnel administered the tests at school on regular school days. Each child was individually administered the four intelligence tests (IDS, RIAS, SON-R 6-40, WISC-IV) in counterbalanced order. Three appointments were required, each about 2 h, including breaks (one test session for the IDS, one test session for the WISC-IV, and one test session for the RIAS and SON-R 6-40). The sample sizes for each intelligence test vary somewhat, as a few children could attend only two testing appointments ( $n_{IDS} = 103$ ,  $n_{RIAS} = 102$ ,  $n_{SON-R\ 6-40} = 101$ ,  $n_{WISC-IV} = 103$ ). After the study was completed, the parents received a written report on their child's performance in each intelligence test. Three years later, parents were contacted again and asked to provide information about their child's school grades. Two families had moved and could not be reached; in all, 54 parents returned the requested information (resulting in a response rate of 53%).

### Data Analyses

Regression analyses using SPSS 21 were conducted. For all analyses, standardized intelligence test scores ( $M = 100$ ,  $SD = 15$ ) were used. To analyze the predictive validity of each intelligence test, separate regression analyses for each predictor (i.e., general intelligence indices) and outcome variable (i.e., child's school grades) were conducted. All variables entered into the regression analyses were z-standardized. In the following analyses, we controlled for variables that showed correlations with the outcome variables to some extent (i.e., sex, age; see Table 1). Standardized regression coefficients (.10 as small, .30 as medium, and .50 as large; after Cohen, 1988) with corresponding  $p$  values are presented.

---

Please insert Table 1 about here

---

### Results

Correlations among all variables are displayed in Table 1. The general intelligence indices of all four tests correlated highly with each other ( $r = .62$  to  $.80$ ,  $p < .001$ ).

Table 2 gives an overview of the descriptive statistics of the current sample. The mean scores of the intelligence tests were somewhat higher than in the standardization samples ( $M = 100$ ), and the standard deviations were somewhat lower than in the standardization samples

## INTELLIGENCE AND SCHOOL GRADES

( $SD = 15$ ). The range of school grades is narrow (4 to 6) and reflects grades in the passing range.

---

Please insert Table 2 about here

---

Table 3 presents the results of the prediction of longitudinal school grades by intelligence while controlling for children's sex and age.

When school grades were averaged across subjects, the general intelligence indices of all four intelligence tests predicted scholastic achievement ( $\beta = .259$  to  $.366$ ,  $p < .05$ ) with small to moderate effect sizes.

Regarding mathematics, results show that IDS ( $\beta = .341$ ,  $p < .05$ ) was a moderate and SON-R 6-40 ( $\beta = .273$ ,  $p < .05$ ) was a weak predictor. The other general intelligence indices did not predict school grades in mathematics. Regarding language, IDS ( $\beta = .307$ ,  $p < .05$ ) and RIAS ( $\beta = .322$ ,  $p < .01$ ) moderately predicted school grades. The general intelligence indices of SON-R 6-40 and WISC-IV did not predict school grades in language.

---

Please insert Table 3 about here

---

### Discussion

Our main goal was to assess the longitudinal predictive validity of four intelligence tests currently used in German-speaking countries for children's school grades in mathematics, language, and averaged across subjects.

The general intelligence indices of all four intelligence tests showed moderate predictive validity for averaged school grades three years after intelligence assessment, which is in line with previous studies, showing that intelligence is a positive predictor of scholastic achievement (Deary et al., 2007; Gut et al., 2012, 2013; Roth et al., 2015). Therefore, our results support the use of the general intelligence indices of the IDS, RIAS, SON-R 6-40, and WISC-IV in order to make predictions of a child's averaged school grades.

Regarding the prediction of mathematics, IDS and SON-R 6-40 were weakly to moderately associated with school grades. IDS general intelligence includes four out of seven subtests that tax phonological and visual-spatial working memory (i.e., selective attention, phonological memory, visual-spatial memory, auditory memory). Previous research revealed that both aspects of working memory are associated with mathematics (e.g., Dehn, 2008; Raghubar, Barnes, and Hecht, 2010). SON-R 6-40, in turn, assesses intelligence through subtests measuring primarily visual-spatial abilities. Previous literature found visual-spatial abilities to be moderately associated with mathematics (e.g., Verdine, Golinkoff, Hirsh-Pasek, Newcombe, Filipowicz, and Chang, 2014; Wai, Lubinski, and Benbow, 2009). The other general intelligence indices did not significantly predict school grades in mathematics, although the small effect sizes were positive; however, this contradicts the results of previous studies that found general intelligence to be a moderate to strong predictor of school grades in mathematics (Roth et al., 2015). The RIAS does not include subtests assessing working memory, while the WISC-IV includes perceptual reasoning and phonological working memory as two out of four specific indices, and thus measure visual-spatial abilities and working memory to a lesser extent. This might have weakened the relation between these general intelligence indices and school grades in mathematics, as visual-spatial abilities (e.g., Verdine et al., 2014; Wai et al., 2009) and working memory (Dehn, 2008; Raghubar et al., 2010) were found to be predictors of mathematics. Thus, in the IDS, phonological and visual-spatial working memory capacity, and in the SON-R 6-40, visual-spatial abilities are considered more important parts of intelligence compared to the other intelligence tests. Therefore, it might be plausible that in particular IDS and SON-R 6-40 were significantly associated with school grades in mathematics.



## INTELLIGENCE AND SCHOOL GRADES

Regarding the prediction of language, the general intelligence indices of the IDS and RIAS were moderately associated with school grades in language. The association between the IDS and language is in line with studies revealing a moderate to strong relationship between working memory and language (e.g., Dehn, 2008). The association between the RIAS and language might be explained through the high requirements of verbal abilities and verbal reasoning in two out of four RIAS subtests. The other general intelligence indices showed no significant associations with language, although the small effect sizes were positive; however, this result contradicts the findings of previous studies that found general intelligence to be a moderate to strong predictor of school grades in language (Roth et al., 2015). In contrast to the RIAS, the SON-R 6-40 focuses only on nonverbal intelligence. Furthermore and in contrast to the IDS, the SON-R 6-40 does not include subtests taxing working memory, which is considered as being associated with language (e.g., Dehn, 2008). The WISC-IV includes a specific index for verbal comprehension and working memory. However, it might be possible that these two out of four specific indices were not sufficient to significantly explain variance in school grades in language in the present study.

It is notable that the effect sizes in the present study were in the small to moderate range and thus somewhat lower than we expected based on the meta-analytical results by Roth et al. (2015). However, in a single study, the expected effects may be smaller, as seen, for example, in Gut et al. (2012, 2013), for several reasons. In the present study, for instance, the analyzed sample showed slightly higher general intelligence scores than the population and had a narrow range in school grades, which were all in the passing range. This might have led to range restrictions, which may have weakened the correlations between intelligence and school grades in the present study (Roth et al., 2015; Sternberg et al., 2001). Also, the present study analyzed the predictive validity of intelligence tests on school grades for elementary school children. According to Roth et al. (2015), lower intelligence scores in elementary school might be better compensated with practice than in higher-grade levels, which may also have led to smaller effect sizes in the present study.

In sum, our results indicate that the general intelligence indices of the German versions of the IDS, RIAS, SON-R 6-40, and the WISC-IV moderately predicted averaged school grades over three years. Furthermore, the IDS was moderately and positively associated with longitudinal mathematics and language school grades, while SON-R 6-40 was a small to moderate predictor of mathematics school grades and RIAS was a moderate predictor of language school grades. Thus, our results provide evidence for predictive validity of these intelligence tests (Neukrug and Fawcett, 2015).

The current study has strengths and limitations. It is a strength that we analyzed four intelligence tests currently used in German-speaking countries, as there is a paucity of information regarding their predictive validity for school grades. Furthermore, we assessed intelligence three years prior to school grades being inquired and could therefore analyze their predictive validity longitudinally. This is especially relevant when practitioners use intelligence scores in order to predict future scholastic achievement. Finally, we measured a child's scholastic achievement in mathematics and language using school grades, which reflect a child's performance and effort over an extended period of time and which are crucial for further scholastic and occupational qualifications (Roth et al., 2015). However, the association between intelligence and scholastic achievement may vary with different operationalization of scholastic achievement. In order to avoid potential errors in parental reports, future studies analyzing currently used intelligence tests in German-speaking countries might also consider achievement tests, which measure specific scholastic abilities at a specific point in time (Rost, 2009), as well as official school records obtained directly from schools. Moreover, as school grades are considered as indicators of achievement over a longer time period, they may be influenced not only by intelligence but also by other constructs (Rost, 2009). Therefore, future studies might also consider noncognitive factors that

## INTELLIGENCE AND SCHOOL GRADES

additionally predict scholastic achievement, such as school engagement (Reyes, Brackett, Rivers, White, and Salovey, 2012), motivation (Steinmayr and Spinath, 2009), self-control (Duckworth, Quinn, and Tsukayama, 2012), personality (Poropat, 2009), and social-emotional competencies (Gut et al., 2012).

Furthermore, the current study had a high drop-out rate (although comparable to that of the studies conducted by Gut et al., 2012, 2013) for the longitudinal information on the child's school grades. This led to a small sample size at Study wave 2. The statistical power of the present study was sufficient to detect expected moderate associations, but there was not enough statistical power to detect weak associations between intelligence and school grades, as discussed above. Furthermore, the present study examined typically developing children enrolled in regular school with slightly higher intelligence. Thus, the conclusions based on the current study cannot be generalized to children with special needs or with different intelligence levels. To examine the predictive validity of the present intelligence tests, future studies are required with larger sample sizes and including children with different levels of intelligence (e.g., children with intellectual disabilities) or special needs as seen in the studies of Canivez, Watkins, James, Good, and James (2014), Mayes and Calhoun (2007), as well as Nelson and Canivez (2012). Because of these limitations, findings from the current study have to be considered as preliminary results.

In conclusion, general intelligence measured with the German version of the IDS, RIAS, SON-R 6-40, and WISC-IV was a positive predictor of averaged school grades in the current longitudinal study. These results support the use of the four intelligence tests for issues raised in psychological practice and reveal their predictive validity on longitudinal scholastic achievement in typically developing school-aged children with slightly higher intelligence. Furthermore, the IDS could predict both school grades in mathematics and language, while the SON-R 6-40 could predict school grades in mathematics and the RIAS in language. These results suggest that school grades in mathematics and language can be predicted by intelligence tests depending on their composition of subtests (e.g., working memory, verbal abilities, visual-spatial abilities). Thus, in psychological practice, examiners have to consider the variety of subtests included in a particular subtest when making specific predictions of mathematics and language. More studies analyzing larger samples as well as children with different levels of intelligence or special needs are required to replicate and generalize the findings of the current study.

### Author Contribution Statement

JG contributed to the study design, acquisition, analysis, and interpretation of data. Drafted and revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

PH contributed to the study design, acquisition, analysis, and interpretation of data. Drafted and revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FS contributed to the study design and acquisition of data. Revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

AG contributed to the study design and interpretation of data. Revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that

## INTELLIGENCE AND SCHOOL GRADES

questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

### **Funding**

There was no funding source for this study.

### **Acknowledgements**

We thank Anita Todd and Laura Wiles for proofreading.

## INTELLIGENCE AND SCHOOL GRADES

## References

- Binet, A., and Simon, T. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Anée Psychologique*, 11, 191-336.
- Canivez, G. L., Watkins, M. W., James, T., Good, R., and James, K. (2014). Incremental validity of WISC-IV<sup>UK</sup> factor index scores with a referred Irish sample: Predicting performance on the WIAT-II<sup>UK</sup>. *British Journal of Educational Psychology*, 84, 667-684. doi:10.1111/bjep.12056
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam, Netherlands: North-Holland.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Conway, A. R. A., Getz, S. J., Macnamara, B., and Engel de Abreu, P. M. J. (2011). Working memory and intelligence. In R. J. Sternberg and S. B. Kaufman (eds.). *Cambridge handbook of intelligence*. (pp. 394-418). New York, NY: Cambridge University Press.
- Deary, I. J., Strand, S., Smith, P., and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21. doi:10.1016/j.intell.2006.02.001
- Dehn, M. J. (2008). *Working Memory and Academic Learning. Assessment and Intervention*. Hoboken, NJ: Wiley.
- Duckworth, A. L., Quinn, P. D., and Tsukayama, E. (2012). What no child left behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104, 439-451. doi:10.1037/a0026280
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-179. doi:10.3758/BF03193146
- Glutting, J. J., Watkins, M. W., Konold, T. R., and McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, 40, 103-114. doi:10.1177/00224669060400020101
- Grob, A., Meyer, C. S., and Hagmann-von Arx, P. (2013). *Intelligence and Development Scales (IDS)* (2nd ed.). Bern, Switzerland: Hans Huber.
- Gut, J., Reimann, G., and Grob, A. (2012). Cognitive, language, mathematical, and socioemotional skills as predictors of subsequent school performance. *Zeitschrift für Pädagogische Psychologie*, 26, 213-200. doi:10.1024/1010-0652/a000070
- Gut, J., Reimann, G., and Grob, A. (2013). A contextualized view on long-term predictors of academic performance. *Journal of Educational Psychology*, 105, 436-443. doi:10.1037/a0031503
- Hagmann-von Arx, P., Gauck, L., and Grob, A. (2015). Intelligence assessment. In K. Seifried, S. Drewes, and M. Hasselhorn (Eds.), *Handbook of school psychology* (pp. 128-139). Stuttgart, Germany: Kohlhammer.
- Hagmann-von Arx, P., and Grob, A. (2014). *Reynolds Intellectual Assessment Scales (RIAS)*. Bern, Switzerland: Hans Huber.
- Heller, K. A. & Perleth, C. (2000). *German Cognitive Ability Test – 4-12 - Revision*. Göttingen, Germany: Hogrefe.
- Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford, England: Oxford University Press.
- Mayes, S. D., and Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children-Third and -Fourth Edition predictors of academic achievement in children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly*, 22, 234-249. doi:10.1037/1045-3830.22.2.234

## INTELLIGENCE AND SCHOOL GRADES

- Nelson, J. M., and Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Assessment Scales (RIAS) with a clinical sample. *Psychological Assessment*, 24, 129-140. doi:10.1037/a0024878
- Neukrug, E. S., and Fawcett, R. C. (2015). *The essentials of testing and assessment: A practical guide for counselors, social workers, and psychologists* (3rd ed.). Stamford, CT: Cengage Learning.
- Petermann, F., and Petermann, U. (2011). *WISC-IV: Wechsler Intelligence Scale for Children—Fourth Edition* (German version). Frankfurt, Germany: Pearson Assessment.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135, 322-338. doi:10.1037/a0014996
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and individual differences*, 20, 110-122. doi:10.1016/j.lindif.2009.10.005
- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., and Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, 104, 700-712. doi:10.1037/a0027268
- Reynolds, C. R., and Kamphaus, R. W. (2003). *RIAS, Reynolds Intellectual Assessment Scales*. Lutz, FL: PAR.
- Rost, D. H. (2009). *Intelligence: Facts and myths*. Weinheim, Germany: Beltz.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., and Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118-137. doi:10.1016/j.intell.2015.09.002
- Steinmayr, R., and Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, 19, 80-90. doi:10.1016/j.lindif.2008.05.004
- Sternberg, R. J., Grigorenko, E., and Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, 47, 1-41. doi:10.1353/mpq.2001.0005
- Swiss Media Institute for Education and Culture. (2016). *Compulsory education*. Retrieved from <http://swisseducation.educa.ch/en/compulsory-education-0>
- Swiss Federal Statistical Office. (2016). *A portrait of Switzerland. Results from the 2010-2014 censuses*. Retrieved from <https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungsstand-kompetenzen/bevoelkerung.html>
- Tellegen, P. J., Laros, J. A., and Petermann, F. (2012). *Snijders-Oomen Nonverbal Intelligence Test (SON-R 6-40)*. Göttingen, Germany: Hogrefe.
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., and Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development*, 85, 1062-1076. doi:10.1111/cdev.12165
- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101, 817-835. doi: 10.1037/a0016127
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) administration and scoring manual*. San Antonio, TX: The Psychological Association.
- Weiss, R. H. (2006). *Culture Fair Intelligence Test Scale 2—Revision (CFT 20-R)*. Göttingen, Germany: Hogrefe.

## INTELLIGENCE AND SCHOOL GRADES

Table 1  
*Correlations Among Control Variables, the Intelligence Scores of the IDS, RIAS, SON-R 6-40, WISC-IV, and School Grades*

Variable	N	1	2	3	4	5	6	7	8	9	10
1 Sex (0 = males, 1 = females)	54	1									
2 Age at Study wave 1	54	-.09	1								
3 Age at Study wave 2	54	-.09	.79***	1							
4 Parental education (0 = non-tertiary, 1 = tertiary)	54	-.02	-.14	-.06	1						
5 IDS	54	.15	-.39**	-.25	.13	1					
6 RIAS	54	-.01	-.32*	-.12	.18	.78***	1				
7 SON-R 6-40	52	.06	-.18	-.17	.08	.75***	.65***	1			
8 WISC-IV	54	.08	-.30*	-.08	-.01	.80***	.70***	.62***	1		
9 Averaged school grade	54	.24	-.30*	-.37**	.10	.46**	.35**	.34*	.30*	1	
10 Mathematics school grade	54	.14	-.25	-.28*	.13	.40**	.27	.32*	.24	.93***	1
11 Language school grade	53	.32*	-.30*	-.41**	.06	.43**	.37**	.29*	.30*	.91***	.68***

*Notes.* Correlations were calculated for individuals who participated at Study wave 2. IDS = Intelligence and Development Scales; RIAS = Reynolds Intellectual Assessment Scales; SON-R 6-40 = Snijders-Oomen Nonverbal Intelligence Test Revised; WISC-IV = Wechsler Intelligence Scale for Children – Fourth Edition.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

## INTELLIGENCE AND SCHOOL GRADES

Table 2  
*Descriptive Statistics for Control Variables, the Intelligence Scores of the IDS, RIAS, SON-R 6-40, WISC-IV, and School Grades in Study Waves 1 and 2*

Variable	Study wave 1 (N = 103)					Study wave 2 (N = 54)				
	Mean (SD)	N (%)	Range	Skewness	Kurtosis	Mean (SD)	N (%)	Range	Skewness	Kurtosis
Sex										
Female		54 (52)					28 (52)			
Male		49 (48)					26 (48)			
Age at Study wave 1 (years)	9.18 (0.93)		6.71–11.18	-0.19	0.26	9.04 (0.96)		6.71–11.18	-0.08	-0.14
Age at Study wave 2 (years)						11.77 (0.80)		10.25–13.60	0.43	-0.51
Parental education										
Non-tertiary education		74 (72)					40 (74)			
Tertiary education		29 (28)					14 (26)			
IDS	103.54 (10.64)		78–134	0.15	0.41	105.19 (10.74)		79–134	0.45	0.70
RIAS	101.45 (8.99)		81–128	0.29	0.39	103.65 (8.23)		87–128	0.84	0.67
SON-R 6-40	103.27 (11.32)		79–139	0.38	0.50	103.67 (11.55)		82–139	0.62	0.61
WISC-IV	104.88 (11.91)		73–133	0.21	0.07	107.39 (10.58)		84–133	0.27	0.28
Averaged school grade						5.16 (0.42)		4–6	-0.39	-0.20
Mathematics school grade						5.15 (0.48)		4–6	-0.25	-0.71
Language school grade						5.16 (0.43)		4–6	-0.31	-0.08

*Notes.* IDS = Intelligence and Development Scales; RIAS = Reynolds Intellectual Assessment Scales; SON-R 6-40= Snijders-Oomen Nonverbal Intelligence Test Revised; WISC-IV = Wechsler Intelligence Scale for Children – Fourth Edition.

## INTELLIGENCE AND SCHOOL GRADES

Table 3  
*Regression Analyses for the Intelligence Scores of the IDS, RIAS, SON-R 6-40, and WISC-IV Predicting Longitudinal School Grades*

Variables	Averaged school grade			Mathematics			Language		
	N	$\beta$	$R^2$	N	$\beta$	$R^2$	N	$\beta$	$R^2$
Step 1			180**			092			251**
Sex (0 = males, 1 = females)	54	.205		54	.120		53	.286*	
Age	54	-.353**		54	-.269		53	-.387**	
Step 2 <sup>a</sup>									
IDS	54	.366**	.303***	54	.341*	.200*	53	.307*	.337***
RIAS	54	.315*	.277**	54	.238	.148*	53	.322**	.353***
SON-R 6-40	52	.274*	.252***	52	.273*	.165*	51	.199	.289**
WISC-IV	54	.259*	.246**	54	.212	.137	53	.236	.306***

Notes. IDS = Intelligence and Development Scales; RIAS = Reynolds Intellectual Assessment Scales; SON-R 6-40= Snijders-Oomen Nonverbal Intelligence Test Revised; WISC-IV = Wechsler Intelligence Scale for Children – Fourth Edition.

<sup>a</sup> Controlled for variables in step 1.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .



**APPENDIX D: Study 4**

Hagmann-von Arx, P., Gygi, J. T., Weidmann, R., & Grob, A. (2016). Testing relations of crystallized and fluid intelligence and the incremental predictive validity of conscientiousness and its facets on career success in a small sample of German and Swiss workers. *Frontiers in Psychology*, 7, 500.



# Testing Relations of Crystallized and Fluid Intelligence and the Incremental Predictive Validity of Conscientiousness and Its Facets on Career Success in a Small Sample of German and Swiss Workers

Priska Hagmann-von Arx\*, Jasmin T. Gygi, Rebekka Weidmann and Alexander Grob

Department of Psychology, University of Basel, Basel, Switzerland

## OPEN ACCESS

### Edited by:

Tim Bogg,  
Wayne State University, USA

### Reviewed by:

Katherine S. Corker,  
Kenyon College, USA  
Jason L. Huang,  
Michigan State University, USA  
Eric Mayor,  
University of Neuchâtel, Switzerland

### \*Correspondence:

Priska Hagmann-von Arx  
priska.hagmann@unibas.ch

### Specialty section:

This article was submitted to  
Personality and Social Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 October 2015

**Accepted:** 23 March 2016

**Published:** 14 April 2016

### Citation:

Hagmann-von Arx P, Gygi JT,  
Weidmann R and Grob A (2016)  
Testing Relations of Crystallized and  
Fluid Intelligence and the Incremental  
Predictive Validity of  
Conscientiousness and Its Facets on  
Career Success in a Small Sample of  
German and Swiss Workers.  
Front. Psychol. 7:500.  
doi: 10.3389/fpsyg.2016.00500

This study examined the relation of fluid and crystallized intelligence with extrinsic (occupational skill level, income) and intrinsic (job satisfaction) career success as well as the incremental predictive validity of conscientiousness and its facets. Participants ( $N = 121$ ) completed the Reynolds Intellectual Assessment Scales (RIAS), the Revised NEO Personality Inventory (NEO-PI-R), and reported their occupational skill level, income, and job satisfaction. Results revealed that crystallized intelligence was positively related to occupational skill level, but not to income. The association of crystallized intelligence and job satisfaction was negative and stronger for the lowest occupational skill level, whereas it was non-significant for higher levels. Fluid intelligence showed no association with career success. Beyond intelligence, conscientiousness and its facet self-discipline were associated with income, whereas conscientiousness and its facets competence and achievement striving were associated with job satisfaction. The results are discussed in terms of their implications for the assessment process as well as for future research to adequately predict career success.

**Keywords:** intelligence, conscientiousness, career success, occupational skill level, income, job satisfaction

## INTRODUCTION

Career success is of importance to individuals as well as to organizations, because it has the capacity to contribute to organizational success (Judge et al., 1999). Career success can be defined as the positive work and psychological outcomes, which have resulted from one's work experiences (Seibert and Kraimer, 2001). It can be structured into extrinsic, objective indicators such as occupational skill level and income as well as intrinsic, subjective indicators such as job satisfaction (Judge et al., 1999; Ng et al., 2005). In addition, the term job performance refers to the performance within one's occupation (Schmidt and Hunter, 2004).

Research in psychology has shown great interest in examining predictors of career success. First, intelligence, the "can do" of a person (Guion and Gottier, 1965, p. 151), has been found to be the strongest predictor of career success (Schmidt and Hunter, 2004). Second, the personality trait conscientiousness, the so-called "will do" (Gottfredson, 2002, p. 37) of an individual, has

also been established as an important trait for career success (Hurtz and Donovan, 2000; Sackett and Walmsley, 2014). Both predictors are embedded in a recent integrative theoretical model proposed by Schmidt (2014), suggesting that intelligence, foremost crystallized intelligence (that originally required fluid intelligence for its development), as well as conscientiousness exert direct influence on an adult's occupational achievement. Studies have also shown that conscientiousness demonstrates incremental validity for determining career success and job performance above and beyond intelligence (Schmidt and Hunter, 1998; Judge et al., 1999; Avis et al., 2002). This study extends current research in not only examining the incremental validity of conscientiousness in the prediction of career success beyond fluid and crystallized intelligence but also by analyzing the trait conscientiousness as a whole as well as in its separate narrow facets.

Intelligence can be considered as equivalent to the general factor *g*, superordinate to all cognitive abilities. Intelligence models (Horn and Cattell, 1966) suggest that *g* can be divided into two separate general factors, namely fluid *g* (reasoning and problem solving, independent of acquired knowledge) and crystallized *g* (accumulated information and verbal skills).

Meta-analytical results show that general intelligence is the best single predictor for extrinsic career success—such as occupational level attained—with moderate correlations around 0.50 (Schmidt and Hunter, 2004). Further, meta-analytic results show smaller associations between general intelligence and income ranging from 0.20 to 0.27 (Ng et al., 2005; Strenze, 2007).

General intelligence can also be linked to intrinsic career success such as job satisfaction, although this association seems to vary depending on job complexity or level of occupation. For instance, empirical results show that general intelligence has a small positive effect on job satisfaction when job complexity is high, whereas it has a small negative effect when job complexity is low (Ganzach and Fried, 2012). Likewise, another study found that the association between intelligence and career satisfaction was significantly negative for lower occupational levels such as hourly employees ( $r = -0.30$ ) but significantly positive for higher occupational levels such as managers ( $r = 0.30$ ; Lounsbury et al., 2004).

Conscientiousness is a dimension of the Big Five model of personality (Costa and McCrae, 1995) and has been shown to be one of the most central personality traits associated with outcomes in the working field (Sackett and Walmsley, 2014). A meta-analytic study reported conscientiousness was linked to work performance across several assessed occupational groups with estimated true correlations above 0.20 (Hurtz and Donovan, 2000). Furthermore, meta-analytic results suggest that conscientiousness is weakly associated with income ( $r = 0.07$ ; Ng et al., 2005) and intrinsic career success such as job or career satisfaction with correlations ranging from 0.14 to 0.26 (Judge et al., 2002; Ng et al., 2005).

As conscientiousness is not a unidimensional construct (Costa and McCrae, 1995), it can be assumed that facets of conscientiousness may each contribute a different amount of variance to career success and job performance. For instance, a meta-analysis (Dudley et al., 2006) showed that

the conscientiousness facets achievement, dependability, order, and cautiousness differentially increased the explained variance of overall job performance by 1–24% over and above global conscientiousness. In these analyses, occupational type (sales personnel, customer service representatives, managers, skilled, and semiskilled workers) served as a moderator of the relation between conscientiousness and job performance. For example, in managers there was a negative relation between order (i.e., being well-organized and methodical) and job performance, while in skilled and semiskilled workers the association was positive. Furthermore, a recent meta-analysis (Judge et al., 2013) tested a hierarchical framework in which the trait conscientiousness comprised two lower order traits, industriousness and orderliness, proposed by DeYoung et al. (2007). Both lower order traits encompassed three facets of conscientiousness introduced by Costa and McCrae (1995). Industriousness included the facets achievement striving, competence, and self-discipline, whereas orderliness comprised the facets deliberation, dutifulness, and order. Results showed that the broad trait conscientiousness as well as the two lower order traits industriousness and orderliness were related to job performance with correlations ranging from 0.21 to 0.26. The facets contributed differently to the prediction of job performance with achievement striving, dutifulness, and self-discipline showing the highest associations ranging from 0.19 to 0.24. Although, the effect sizes were modest, the authors concluded that the assessment of lower order traits improved criterion-related validity of job performance over that of the broad trait (Judge et al., 2013).

Research results have revealed that intelligence and conscientiousness independently predict career success. Further, studies have also focused on the question of whether conscientiousness provides incremental validity beyond intelligence in predicting career success and job performance (Schmidt and Hunter, 1998; Judge et al., 1999; Avis et al., 2002). For example, Schmidt and Hunter (1998) reported in a meta-analysis that conscientiousness predicted job performance above and beyond cognitive ability with 10% incremental variance explained. However, we know of no study that has separately considered fluid and crystallized intelligence, as suggested in the theoretical model proposed by Schmidt (2014). Furthermore, we know of no study that has analyzed the incremental validity of facets of conscientiousness on career success beyond fluid and crystallized intelligence.

The current study pursued two objectives. First, we examined the association between intelligence (fluid, crystallized) and extrinsic (occupational skill level, income) and intrinsic (job satisfaction) career success. Based on previous research (e.g., Schmidt, 2014), we expected that intelligence and in particular crystallized intelligence would be a positive predictor of career success. Second, we examined the incremental predictive validity of conscientiousness and its facets in predicting career success. We expected that conscientiousness would be positively related to career success, that conscientiousness would explain incremental predictive variance, and that facets of conscientiousness would contribute differentially to the explanation of career success. Finally, as previous research

suggests that the association between intelligence and job satisfaction may vary depending on occupational skill level, we additionally investigated the moderating role of occupational skill level in this association. In low occupational skill levels we expected that intelligence would be negatively related to career satisfaction, whereas in high occupational skill levels we expected that intelligence would be positively related to career satisfaction. Our study extends previous research by investigating both fluid and crystallized intelligence as well as conscientiousness and its facets in order to clarify their simultaneous associations with career success.

## MATERIALS AND METHODS

### Participants and Procedure

The sample consisted of 121 adults (48 males, 73 females) with an average age of 48.45 years ( $SD = 12.54$  years). The recruitment took place as part of the German standardization and validation of the Reynolds Intellectual Assessment Scales (RIAS; Hagmann-von Arx and Grob, 2014). All participants provided written informed consent to participate. The study and the consent procedure were approved by the Ethics Committee of Basel and the study was performed in accordance with the ethical standards laid down in the Declaration of Helsinki. Participants were from Germany ( $n = 49$ ) and from Switzerland ( $n = 72$ ). At the end of the study, participants received a written report on their test results.

### Measures

Intelligence was assessed using the German version of the RIAS. The RIAS is an individually administered intelligence test for persons between the ages of 3 and 99 years standardized in Germany and Switzerland. It is composed of a two-subtest measure of non-verbal intelligence and a two-subtest measure of verbal intelligence, both of which were developed to closely match the constructs of fluid and crystallized intelligence. An overall intelligence score can be calculated from the sum of the T scores of the four subtests. In the current sample, the internal consistency for the overall intelligence score ( $\alpha = 0.90$ ) as well as for non-verbal/fluid ( $\alpha = 0.83$ ) and verbal/crystallized ( $\alpha = 0.91$ ) intelligence was high.

Conscientiousness and its facets (Competence, Order, Dutifulness, Achievement Striving, Self-Discipline, Deliberation) were assessed with the German version of the NEO Personality Inventory-Revised (NEO-PI-R; Ostendorf and Angleitner, 2004). The NEO-PI-R is a self-report inventory containing 240 items, grouped into 30 facet scales, which are hierarchically organized under the five major dimensions of personality (i.e., Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness). Each facet contains eight items. Responses are made on a five-point scale (1 = *strongly disagree*; 5 = *strongly agree*). In the current sample, the internal consistency for the composite score of conscientiousness was high ( $\alpha = 0.82$ ). Facet reliabilities were moderate to high with  $\alpha = 0.65$ – $0.81$ .

Job satisfaction was assessed using a short German self-report survey (Neuberger and Allerbeck, 1978) containing eight items,

which were rated on a five-point scale (1 = *does not apply at all*; 5 = *applies completely*). Example items are “I really enjoy my work.” and “I am always in a rut with my work; nothing can be done about it” (reverse-scored). In the current sample, a moderate Cronbach's alpha value of 0.76 was recovered.

Personal income was assessed with one item: “How high is your gross income per year?” which was answered by  $n = 76$  (63%) subjects. To control for cross-country differences in income, participants' income was divided by their country's most recent purchasing power parity to reflect participants' personal purchasing power within their country. We log-transformed the income according to suggestions by Kahneman and Deaton (2010).

Occupational skill level was calculated on the basis of the participants' profession. This information was encoded according to the four skill levels (1 = *unskilled*; 4 = *highly skilled*) distinguished in the International Standard Classification of Occupations (ISCO-08; International Labour Organization, 2008). Descriptive statistics for demographic variables, intelligence, conscientiousness, and career success including mean, standard deviation, range, skew, and kurtosis are shown in Table 1.

### Statistical Procedure

All analyses were carried out using SPSS 22.0. As the sample size was rather small and distributions of some of the variables showed deviations from normality, we used bootstrap procedures (Efron, 1979; Chernick, 2008) for all analyses to construct bias-corrected 95% confidence intervals (BC 95%-CI) based on 5000 random samples. When the confidence interval did not include zero, an effect was considered as significant.

Two sets of simultaneous regression analyses were conducted for each of the career success outcomes. In a first set of analyses, we examined the role of fluid and crystallized intelligence as well as the composite score of conscientiousness in predicting (a) occupational skill level, (b) logged income, and (c) job

**TABLE 1 | Descriptive statistics for demographic variables, intelligence, conscientiousness, and career success.**

Variable	N	M (%)	SD	Range	Skew	Kurtosis
Sex (men)	121	(40)	–	–	–	–
Age (years)	121	48.45	12.54	21.67–77.33	–0.275	–0.356
Fluid intelligence	121	101.80	10.66	61–123	–0.363	0.728
Crystallized intelligence	121	103.26	11.19	66–123	–0.806	0.402
Conscientiousness	121	127.65	18.77	77–184	0.213	0.962
Competence	121	22.43	3.63	11–32	0.045	0.814
Order	121	19.69	4.69	4–30	–0.564	0.827
Dutifulness	121	23.82	3.69	15–31	–0.310	–0.069
Achievement striving	121	19.29	4.26	8–31	–0.073	0.298
Self-discipline	121	19.98	5.23	4–32	–0.436	0.176
Deliberation	121	17.39	4.40	6–28	–0.271	–0.544
Occupational skill level	120	3.20	0.89	1–4	–0.627	–0.885
Logged income	76	4.54	0.28	3.66–5.08	0.632	0.653
Job satisfaction	118	2.70	0.64	1–4	–0.740	0.749

satisfaction, controlling for age and sex. The second set of analyses included fluid and crystallized intelligence as well as the six facets of conscientiousness predicting (a) occupational skill level, (b) logged income, and (c) job satisfaction, controlling for age and sex.

Regarding the association between intelligence and job satisfaction, two sets of moderated regression analyses were conducted following the procedure proposed by Aiken and West (1991) and Cohen et al. (2013) to investigate whether occupational skill level acts as a moderator of this association. Occupational skill levels 1 and 2 were combined because of a small case number for skill level 1. Thus, occupational skill level was a three-group categorical variable (low, middle, and high occupational skill level) and was dummy coded for inclusion in the regression equations (Tabachnik and Fidell, 2007). In a first set of analyses, control variables, fluid and crystallized intelligence, the composite score of conscientiousness, occupational skill level, and the interaction terms between occupational skill level and intelligence variables were entered into the regression equation predicting job satisfaction. In a second set of analyses, control variables, fluid and crystallized intelligence, the six facets of conscientiousness, occupational skill level, and the interaction terms between occupational skill level and intelligence variables were entered into the regression equation predicting job satisfaction. Variables included in the interaction term were centered. If a significant interaction was identified, indicating a moderation effect, then the interaction was graphed by computing predicted values of job satisfaction separately for each occupational skill level at low ( $-1$  SD) and high ( $+1$  SD) values of intelligence. Analyses of simple slopes were conducted to evaluate whether the slopes of the independent variables were significantly different from zero in each occupational skill level (Cohen et al., 2013). For all variables, z-standardized scores were used such that the reported unstandardized estimates can be interpreted as standardized regression coefficients.

## RESULTS

As shown in Table 2, in the present sample, sex was correlated to logged income ( $r = -0.281$ ,  $p = 0.014$ ), such that men reported a higher income than women. The composite score of conscientiousness was strongly positively correlated to its six facets ( $r = 0.440$ – $0.771$ ,  $p < 0.001$ ). To avoid problems of multicollinearity, the composite score (model 1) and the six facets of conscientiousness (model 2) were entered in separate regression models.

The results of the simultaneous regression analyses are shown in Table 3. In model 1 (including the composite score of conscientiousness) crystallized intelligence was significantly related to occupational skill level (Estimate = 0.306,  $SE = 0.112$ , BC 95%-CI = [0.082, 0.514]), whereas fluid intelligence showed no significant association with occupational skill level (Estimate = 0.015,  $SE = 0.097$ , BC 95%-CI = [-0.170, 0.199]). Neither intelligence factors was

related to logged income<sup>1</sup> (crystallized intelligence: Estimate = 0.110,  $SE = 0.135$ , BC 95%-CI = [-0.157, 0.361]; fluid intelligence: Estimate = 0.029,  $SE = 0.130$ , BC 95%-CI = [-0.233, 0.267]), or job satisfaction (crystallized intelligence: Estimate = -0.071,  $SE = 0.099$ , BC 95%-CI = [-0.259, 0.127]; fluid intelligence: Estimate = 0.059,  $SE = 0.089$ , BC 95%-CI = [-0.106, 0.244]). In model 2 (including the six facets of conscientiousness), the results regarding crystallized and fluid intelligence predicting career success were comparable to those in model 1.

The composite score of conscientiousness (model 1) did not explain additional variance in occupational skill level (Estimate = 0.080,  $SE = 0.084$ , BC 95%-CI = [-0.096, 0.254]), although it explained incremental variance in logged income<sup>2</sup> (Estimate = 0.388,  $SE = 0.128$ , BC 95%-CI = [0.128, 0.632]), as well as job satisfaction (Estimate = 0.365,  $SE = 0.088$ , BC 95%-CI = [0.193, 0.549]). In model 2, facets of conscientiousness were not related to occupational skill level. The facet self-discipline (Estimate = 0.584,  $SE = 0.180$ , BC 95%-CI = [0.246, 0.907]) showed a significant positive association with logged income<sup>3</sup>. The facets competence (Estimate = 0.326,  $SE = 0.119$ , BC 95%-CI = [0.095, 0.589]) and achievement striving (Estimate = 0.257,  $SE = 0.108$ , BC 95%-CI = [0.052, 0.478]) were significant positive predictors of job satisfaction (Table 3).

Moderated regression analyses were conducted to examine possible interaction effects of intelligence variables and occupational skill level on job satisfaction. In model 1, these analyses revealed a significant crystallized intelligence  $\times$  occupational skill level interaction for the dummy variable comparing the low with the middle and high occupational skill level (Estimate = -0.255,  $SE = 0.128$ , BC 95%-CI = [-0.512, -0.017]), as shown in Table 4. In model 2, the interaction was no longer significant (Estimate = -0.219,  $SE = 0.143$ , BC 95%-CI = [-0.501, 0.072]). The fluid intelligence  $\times$  occupational skill level interactions for the dummy variable comparing the low with the middle and high occupational skill level as well as for the dummy variable comparing the high with the low and middle occupational skill level did not reach significance, neither in model 1 (low skill level: Estimate = -0.141,  $SE = 0.017$ , BC 95%-CI = [-0.369, 0.073]; high skill level: Estimate = -0.056,  $SE = 0.092$ , BC 95%-CI = [-0.250, 0.110]) nor in model 2 (low skill level: Estimate = -0.151,  $SE = 0.133$ , BC 95%-CI = [-0.400, 0.098]; high skill level: Estimate = -0.051,  $SE = 0.099$ , BC 95%-CI = [-0.260, 0.126]).

To further illuminate the significant interaction effect, we conducted single slope analyses. Controlling for age, sex, non-verbal intelligence, and the composite score of conscientiousness, at the lowest skill level, crystallized intelligence was negatively associated with job satisfaction (Estimate = -0.405,  $SE = 0.166$ , BC 95%-CI = [-0.702, -0.091]) such that lower intelligence was related to higher job satisfaction, whereas in occupational skill level 3 (Estimate = 0.113,  $SE = 0.229$ , BC 95%-CI = [-0.287,

<sup>1, 2</sup>The findings did not change substantially when investigating raw income scores.

<sup>3</sup>When investigating raw income scores the facet competence also reached significance (Estimate = 0.258,  $SE = 0.117$ , BC 95%-CI = [0.017, 0.501]).

**TABLE 2 | Correlations among all variables.**

Variable	N	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Sex (0 = Men, 1 = Women)	121	1												
2 Age (years)	121	0.106	1											
3 Fluid intelligence	121	-0.188	0.087	1										
4 Crystallized intelligence	121	-0.092	-0.060	0.562***	1									
5 Conscientiousness	121	-0.059	0.047	0.075	-0.062	1								
6 Competence	121	0.017	0.147	0.288*	0.184	0.641***	1							
7 Order	121	-0.190	-0.015	-0.127	-0.270*	0.732***	0.153	1						
8 Dutifulness	121	0.012	-0.031	-0.163	-0.203†	0.658***	0.219†	0.473***	1					
9 Achievement striving	121	-0.055	0.006	0.145	0.076	0.686***	0.349**	0.360**	0.338**	1				
10 Self-discipline	121	0.009	-0.002	0.007	-0.088	0.771***	0.247*	0.633***	0.445***	0.446***	1			
11 Deliberation	121	-0.189	-0.121	-0.054	-0.062	0.440***	0.180	0.440***	0.446***	0.279*	0.246*	1		
12 Occupational skill level	120	0.152	0.028	0.206†	0.307**	0.095	0.243*	-0.093	-0.101	0.187	0.019	-0.060	1	
13 Logged income	76	-0.281*	0.159	0.185	0.124	0.362**	0.283*	0.260*	0.062	0.173	0.425***	0.119	0.192†	1
14 Job satisfaction	118	0.043	0.108	0.091	0.007	0.397***	0.383**	0.249*	0.098	0.290*	0.287*	-0.056	0.284*	0.086

† $p < 0.10$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**TABLE 3 | Regression analyses of fluid and crystallized intelligence, conscientiousness (model 1) and its facets (model 2) predicting career success, controlling age and sex.**

Predictor	Occupational skill level (n = 120)			Logged income (n = 76)			Job satisfaction (n = 118)		
	Estimate	SE	BC 95%-CI	Estimate	SE	BC 95%-CI	Estimate	SE	BC 95%-CI
<b>MODEL 1</b>									
Age	0.074	0.084	[-0.089, 0.228]	0.184	0.117	[-0.042, 0.427]	0.112	0.090	[-0.064, 0.293]
Sex	0.181	0.089	[0.002, 0.361]	-0.253	0.100	[-0.445, -0.061]	0.055	0.091	[-0.137, 0.252]
Fluid intelligence	0.015	0.097	[-0.170, 0.199]	0.029	0.130	[-0.233, 0.267]	0.059	0.089	[-0.106, 0.244]
Crystallized intelligence	0.306	0.112	[0.082, 0.514]	0.110	0.135	[-0.157, 0.361]	-0.071	0.099	[-0.259, 0.127]
Conscientiousness	0.080	0.084	[-0.096, 0.254]	0.388	0.128	[0.128, 0.632]	0.365	0.088	[0.193, 0.549]
F of model 1		3.610*			4.566**			4.337**	
R <sup>2</sup> of model 1		0.122			0.246			0.163	
<b>MODEL 2</b>									
Age	0.037	0.084	[-0.124, 0.189]	0.177	0.127	[-0.069, 0.425]	0.056	0.085	[-0.104, 0.200]
Sex	0.175	0.091	[-0.004, 0.353]	-0.292	0.103	[-0.485, -0.102]	0.027	0.093	[-0.157, 0.226]
Fluid intelligence	-0.047	0.099	[-0.239, 0.136]	-0.007	0.135	[-0.263, 0.249]	-0.010	0.099	[-0.192, 0.195]
Crystallized intelligence	0.263	0.111	[0.029, 0.483]	0.085	0.130	[-0.172, 0.315]	-0.118	0.109	[-0.318, 0.098]
Facets of conscientiousness									
Competence	0.213	0.108	[-0.029, 0.437]	0.211	0.130	[-0.059, 0.472]	0.326	0.119	[0.095, 0.589]
Order	0.028	0.121	[-0.210, 0.269]	-0.027	0.150	[-0.301, 0.255]	0.100	0.121	[-0.130, 0.325]
Dutifulness	-0.159	0.119	[-0.398, 0.084]	-0.155	0.132	[-0.425, 0.100]	-0.123	0.097	[-0.310, 0.055]
Achievement striving	0.190	0.119	[-0.029, 0.427]	-0.090	0.151	[-0.372, 0.204]	0.257	0.108	[0.052, 0.478]
Self-discipline	-0.151	0.122	[-0.390, 0.077]	0.584	0.180	[0.246, 0.907]	0.047	0.122	[-0.205, 0.296]
Deliberation	0.000	0.107	[-0.211, 0.189]	0.031	0.126	[-0.206, 288]	-0.178	0.107	[-0.401, 0.022]
F of model 2		2.623**			3.644**			3.721***	
R <sup>2</sup> of model 2		0.194			0.359			0.260	

All variables were z-scored prior to analysis so that all estimates can be interpreted as standardized effects and be directly compared to one another. BC 95%-CI = bias-corrected 95% bootstrap confidence intervals.

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**TABLE 4 | Regression analyses of fluid and crystallized intelligence, conscientiousness (model 1) and its facets (model 2), as well as crystallized intelligence  $\times$  occupational skill level interactions predicting job satisfaction, controlling age and sex.**

Predictor	Job satisfaction ( $n = 118$ )		
	Estimate	SE	BC 95%-CI
<b>MODEL 1</b>			
Age	0.078	0.092	[−0.105, 0.274]
Sex	0.038	0.094	[−0.146, 0.230]
Fluid intelligence	0.069	0.085	[−0.093, 0.241]
Crystallized intelligence	−0.272	0.110	[−0.475, −0.047]
Conscientiousness	0.317	0.089	[−0.137, 0.495]
Occupational skill level			
Low skill level	−0.130	0.118	[−0.364, 0.110]
High skill level	0.198	0.102	[0.003, 0.399]
Crystallized intelligence $\times$ Low skill level	−0.255	0.128	[−0.512, −0.017]
Crystallized intelligence $\times$ High skill level	−0.115	0.118	[−0.356, 0.107]
F of model 1		4.062*	
R <sup>2</sup> of model 1		0.255	
<b>MODEL 2</b>			
Age	0.035	0.084	[−0.132, 0.191]
Sex	0.025	0.099	[−0.179, 0.220]
Fluid intelligence	0.012	0.095	[−0.161, 0.182]
Crystallized intelligence	−0.262	0.129	[−0.501, −0.004]
Facets of conscientiousness			
Competence	0.269	0.128	[0.012, 0.538]
Order	0.063	0.118	[−0.171, 0.317]
Dutifulness	−0.067	0.104	[−0.280, 0.136]
Achievement striving	0.235	0.124	[−0.006, 0.473]
Self-discipline	0.050	0.131	[−0.216, 0.300]
Deliberation	−0.165	0.099	[−0.354, 0.001]
Occupational skill level			
Low skill level	−0.100	0.123	[−0.331, 0.132]
High skill level	0.149	0.110	[−0.068, 0.367]
Crystallized intelligence $\times$ Low skill level	−0.219	0.143	[−0.501, 0.072]
Crystallized intelligence $\times$ High skill level	−0.061	0.132	[−0.310, 0.195]
F of model 2		3.397*	
R <sup>2</sup> of model 2		0.318	

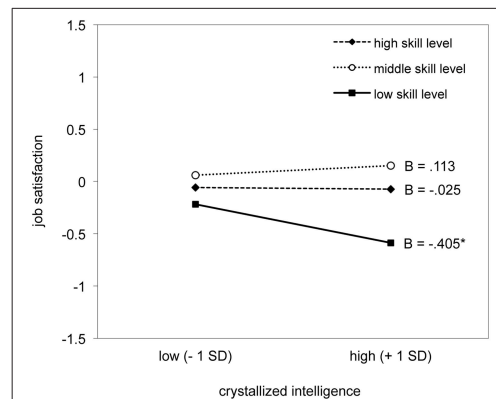
All variables were z-scored prior to analysis so that all estimates can be interpreted as standardized effects and be directly compared to one another. BC 95%-CI = bias-corrected 95% bootstrap confidence intervals.

\* $p < 0.001$ .

0.621]) and 4 (Estimate =  $-0.025$ ,  $SE = 0.148$ , BC 95%-CI =  $[-0.285, 0.280]$ ) there were no associations between crystallized intelligence and job satisfaction, as shown in **Figure 1**.

## DISCUSSION

The current study analyzed the association of intelligence (fluid, crystallized) with extrinsic (occupational skill level, income) and intrinsic (job satisfaction) career success. Furthermore, it examined the incremental predictive validity of conscientiousness and its facets.



**FIGURE 1 | Crystallized intelligence  $\times$  occupational skill level interaction in prediction of job satisfaction.** At the lowest skill level, crystallized intelligence was negatively associated with job satisfaction [Estimate =  $-0.405$ ,  $SE = 0.166$ , BC 95%-CI =  $(-0.702, -0.091)$ ], whereas in the middle skill level [Estimate =  $0.113$ ,  $SE = 0.229$ , BC 95%-CI =  $(-0.287, 0.621)$ ] and high skill level [Estimate =  $-0.025$ ,  $SE = 0.148$ , BC 95%-CI =  $(-0.285, 0.280)$ ] there was no such association. The reported unstandardized estimates can be interpreted as standardized regression coefficients and be directly compared to one another. \* $p < 0.01$ .

Regarding the first hypothesis, whether intelligence is associated with career success, our results revealed that crystallized intelligence is a significant predictor of occupational skill level which is in line with previous studies (Schmidt and Hunter, 2004). Further, this finding corresponds to the integrative theoretical model postulated by Schmidt (2014), which suggests that crystallized intelligence is a major determinant of career success. Crystallized intelligence reflects the knowledge and cognitive skills that a person has acquired through educational and vocational opportunities and thus constitutes expertise. However, the association between crystallized intelligence and occupational skill level was smaller than expected and conflicting with previous research the present study showed no association between fluid intelligence and occupational skill level (Schmidt and Hunter, 2004). Possible explanations may lie in variance restrictions in the present study. The standard deviations for intelligence test scores were lower than in the RIAS standardization sample ( $SD < 15$ ). From previous research it is known that the standard deviation of intelligence test scores increases with decreasing occupational skill level (Schmidt and Hunter, 2004). It may be possible that the small case number of participants with ISCO-08 skill level 1 in the present sample limited the range of intelligence test scores. Thus, restrictions in variance in both intelligence test scores and occupational skill level might have led to attenuated associations between these variables.

Further, intelligence showed no significant relation to income, what is in contrast to previous research (Ng et al., 2005; Strenze, 2007). This result may have emerged because we studied a relatively small sample. Not all participants agreed to answer the open question regarding their income, a recognized



problem in social science research (Yan et al., 2010). Based on meta-analytical results (Ng et al., 2005; Strenze, 2007), the expected effects of the relation between intelligence and income were small. For example, according to Strenze (2007) the average association between intelligence and income is 0.20 with a 95% confidence interval of 0.16–0.23. However, estimated “true correlations” calculated in meta-analyses are usually corrected for range restrictions as well as unreliability of the predictor and outcome. Therefore, in a single study, the expected effect may be even smaller. Regarding the current study, *post-hoc* power analysis using G\*Power (Faul et al., 2007) showed that with a chance above 80% and a 0.05 alpha level we were only able to identify medium-sized effects (Cohen, 1988). Therefore, in the present sample there was not enough statistical power to detect weak associations between intelligence factors and income.

Additionally, we expected that the association between intelligence and job satisfaction varies depending on occupational skill level. This assumption was supported in the present study because results revealed that the association between crystallized intelligence and job satisfaction was moderated by occupational skill level. Crystallized intelligence was a significant negative predictor of job satisfaction in lower occupational skill levels, whereas there was no such association for higher occupational skill levels. The result that in less demanding occupations workers with higher crystallized intelligence report lower satisfaction with their job indicates that workers experiencing cognitive underload are likely to report lower job satisfaction. This is in line with previous research (Lounsbury et al., 2004; Ganzach and Fried, 2012) and might reflect differences in person-work environment fit (Lounsbury et al., 2004). A good person-environment fit can be assumed for workers with high intelligence and high occupational level as they might have the opportunity to develop their potential and therefore be more satisfied with their job. In contrast, a poor person-work environment fit can be assumed for workers with high intelligence and lower occupational level, as they might be restricted in their ability to express their cognitive competence and therefore report a lower job satisfaction. Thus, aligned levels of occupational skill level and intelligence might facilitate job satisfaction.

Regarding the hypothesis, that conscientiousness and its facets explain incremental predictive validity, our results revealed that conscientiousness was significantly positively related to logged income and job satisfaction. These results are in accordance with previous research (Judge et al., 2002; Ng et al., 2005), as well as theoretical assumptions (Schmidt, 2014) and indicate that workers who describe themselves as competent, organized, reliable, self-disciplined, and so on appear to have a higher income and to be more satisfied with their job than workers who describe themselves as being less strong in these characteristics. However, conflicting with previous evidence (Hurtz and Donovan, 2000), in our study conscientiousness was not associated with occupational skill level. From previous research it is known that relations between specific facets of conscientiousness and job performance may differ depending on the investigated profession (Dudley et al., 2006). Our sample

included different professions within an occupational skill level such as lawyers, physiotherapists, and university teachers on ISCO-08 skill level 4. Therefore, it may be possible that this heterogeneity masked potential relations within specific professions and led to non-significant associations between facets of conscientiousness and occupational skill level across the whole sample. Furthermore, as the broad trait conscientiousness is a composite score derived from the six facets it is plausible that the relation between conscientiousness and occupational skill level became non-significant when averaging the non-significant effects of the facets (Dudley et al., 2006; Judge et al., 2013).

More detailed analyses regarding the narrow facets of conscientiousness showed that they contribute differentially to income and job satisfaction. In our sample, self-disciplined people had a higher income, whereas people reporting higher competence and achievement striving were more satisfied with their job. These results are in line with previous research showing that facets of conscientiousness differentially increased the explained variance of career success (Dudley et al., 2006; Judge et al., 2013) and thus underline the importance of acquiring a more fine-grained picture of the association between conscientiousness and career success by including facets of conscientiousness.

In our study we measured conscientiousness and its six facets using the NEO-PI-R, which comprises eight items for each facet. Thus, using only particular conscientiousness facets of the NEO-PI-R provides a more parsimonious assessment than the administration of the whole scale, what leads to more efficient testing. However, it remains to be determined by future studies whether questionnaires shorter than the NEO-PI-R such as the short version of the Big Five Inventory (BFI-K; Rammstedt and John, 2005) show comparable associations with career success than the particular conscientiousness facets of the NEO-PI-R.

Future studies might also shed further light on the association between narrow traits and career success. Regarding narrow conscientiousness traits it might be possible that their association with career success varies depending on the different job stages (Woods et al., 2013). For example, Stewart (1999) showed that in newly hired employees for whom demands were novel and possibly not clearly defined, the narrow trait order (e.g., structuring and organizing the work environment, effectively managing time) was more strongly associated with job performance, whereas in senior employees who have mastered the tasks related to their jobs, the narrow trait achievement (e.g., working hard and persistently to achieve goals) was more strongly associated with job performance. However, in a recent study Ganzach and Pazy (2015) showed that temporal changes in career success are mainly driven by intelligence rather than personality traits. Thus, for future studies it would be of interest to further examine predictors of career success at different stages of people's working lives.

Future research might also investigate potential moderators of the relationship between intelligence, conscientiousness, and career success. Regarding the conscientiousness-career success relationship, the meta-analysis conducted by Shaffer and Postlethwaite (2013) revealed a moderating role of job characteristics such that conscientiousness showed a stronger



association with job performance in highly routinized jobs whereas this association was weaker in jobs requiring higher levels of cognitive ability. It may be possible that this is due to different ways that conscientiousness is measured: in highly routinized jobs conscientiousness may be rated as arriving at work in a timely manner, whereas in more complex jobs conscientiousness may be related to the skill to manage one's own schedules.

Finally, future research might examine underlying processes potentially affecting the association between intelligence, conscientiousness, and career success to further understand the relation between these variables. For instance, Li et al. (2011) suggest that self-esteem could mediate the relationship between intelligence and career success (i.e., leader role occupancy) because individuals' self-esteem rests on their positive evaluations of their competence.

Limitations of our study include the cross-sectional design, which precludes causal inferences and does not allow testing the direction of the effects between intelligence, conscientiousness and career success. Further, conscientiousness and job satisfaction were both measured using self-reports. In order to reduce common method variance (Podsakoff et al., 2003), which may have had an effect on the relationship between conscientiousness and job satisfaction, future studies might use third-party reports to assess conscientiousness. Also, income was a self-reported variable. Although, this is a common method, the sensitivity of this issue may have led to misreporting (Zinn and Würbach, 2016). Therefore, this variable may not accurately reflect true income and thus may be less valid. In addition, the relatively small sample size regarding income does not allow detecting small effects. Future studies may use closed questions with income brackets to reduce missing data. In addition, our sample consisted of higher-qualified collaborators, thus our results cannot be generalized to unskilled workers. Finally, the study investigated subjects with rather heterogeneous professions. Future studies may shed further light on the association of intelligence, conscientiousness and career success by investigating specific types of professions. However, we consider it a strength of the study that both fluid and crystallized intelligence were assessed using a standardized test and that conscientiousness was analyzed as a broad trait as well as in its separate narrow facets. In addition to a more in-depth look at these predictors, we also examined conceptually distinct aspects of career success with occupational skill level and income

representing external aspects and job satisfaction representing an internal aspect of career success. Also, all analyses were controlled for sex, as men reported higher income than women, which is in accordance with the census of Germany (Statistisches Bundesamt, 2014) and Switzerland (Bundesamt für Statistik, 2012).

In conclusion, this study indicates that crystallized intelligence is positively related to occupational skill level and negatively related to job satisfaction in lower occupational skill levels. Hence, examiners may put greater emphasis on crystallized than fluid intelligence when linking intelligence with career success. Beyond intelligence, conscientiousness and its facets—competence, achievement striving, and self-discipline—were differentially associated with income and job satisfaction. To predict career success we therefore propose assessing both intelligence and conscientiousness to combine the “will do” and “can do” aspects of a person.

## AUTHOR CONTRIBUTIONS

PH contributed to the study design, acquisition, analysis, and interpretation of data. Drafted and revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. JG contributed to the acquisition, analysis, and interpretation of data. Drafted and revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. RW contributed to the analysis and interpretation of data. Drafted and revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. AG contributed to the study design and interpretation of data. Revised the manuscript, gave final approval, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## ACKNOWLEDGMENTS

We thank Laura Wiles for proofreading.

## REFERENCES

- Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.
- Avis, J. M., Kudisch, J. D., and Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *J. Bus. Psychol.* 17, 87–105. doi: 10.1023/A:1016200317002
- Bundesamt für Statistik (2012). *Monthly Gross Earnings by Age and Sex 2012*. Available online at: [http://www.bfs.admin.ch/bfs/portal/de/index/themen/03/04/blank/key/lohnstruktur/nach\\_geschlecht.html](http://www.bfs.admin.ch/bfs/portal/de/index/themen/03/04/blank/key/lohnstruktur/nach_geschlecht.html)
- Chernick, M. R. (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers*. Hoboken, NJ: John Wiley & Sons.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Hoboken, NJ: Taylor and Francis.
- Costa, P. T. Jr., and McCrae, R. R. (1995). Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *J. Pers. Assess.* 64, 21–50. doi: 10.1207/s15327752jpa6401\_2
- DeYoung, C. G., Quilty, L. C., and Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *J. Pers. Soc. Psychol.* 93, 880–896. doi: 10.1037/0022-3514.93.5.880

- Dudley, N. M., Orvis, K. A., Lebiecki, J. E., and Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. *J. Appl. Psychol.* 91, 40–57. doi: 10.1037/0021-9010.91.1.40
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7, 1–26. doi: 10.1214/aos/1176344552
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–179. doi: 10.3758/BF03193146
- Ganzach, Y., and Fried, I. (2012). The role of intelligence in the formation of well-being: from job rewards to job satisfaction. *Intelligence* 40, 333–342. doi: 10.1016/j.intell.2012.03.004
- Ganzach, Y., and Pazy, A. (2015). Cognitive versus non-cognitive individual differences and the dynamics of career success. *App Psychol.* 64, 701–726. doi: 10.1111/apps.12038
- Gottfredson, L. S. (2002). Where and why g matters: not a mystery. *Hum. Perform.* 15, 25–46. doi: 10.1207/s15327043hup1501&amp;02\_03
- Guion, R. M., and Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Pers. Psychol.* 18, 135–164. doi: 10.1111/j.1744-6570.1965.tb00273.x
- Hagmann-von Arx, P., and Grob, A. (2014). *Reynolds Intellectual Assessment Scales and Screening (RIAS)TM – German Version*. Bern: Hans Huber.
- Horn, J. L., and Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *J. Educ. Psychol.* 57, 253–270. doi: 10.1037/h0023816
- Hurtz, G. M., and Donovan, J. J. (2000). Personality and job performance: the Big Five revisited. *J. Appl. Psychol.* 85, 869–879. doi: 10.1037/0021-9010.85.6.869
- International Labour Organization (2008). *International Standard Classification of Occupations 2008 (ISCO-08). Structure, Group Definitions and Correspondence Tables*. Geneva: International Labour Office.
- Judge, T. A., Heller, D., and Mount, M. K. (2002). Five-factor model of personality and job satisfaction: a meta-analysis. *J. Appl. Psychol.* 87, 530–541. doi: 10.1037/0021-9010.87.3.530
- Judge, T. A., Higgins, C. A., Thoresen, C. J., and Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Pers. Psychol.* 52, 621–652. doi: 10.1111/j.1744-6570.1999.tb00174.x
- Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., and Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *J. Appl. Psychol.* 98, 875–925. doi: 10.1037/a0033901
- Kahneman, D., and Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16489–16493. doi: 10.1073/pnas.1011492107
- Li, W.-D., Arvey, R. D., and Song, Z. (2011). The influence of general mental ability, self-esteem and family socioeconomic status on leadership role occupancy and leader advancement: the moderating role of gender. *Leadership Quart.* 22, 520–534. doi: 10.1016/j.leaqua.2011.04.009
- Lounsbury, J. W., Gibson, L. W., Steel, R. P., Sundstrom, E. D., and Loveland, J. L. (2004). An investigation of intelligence and personality in relation to career satisfaction. *Pers. Individ. Differ.* 37, 181–189. doi: 10.1016/j.paid.2003.08.010
- Neuberger, O., and Allerbeck, M. (1978). *Measurement and Analysis of Job Satisfaction*. Bern: Hans Huber.
- Ng, T. W., Eby, L. T., Sorensen, K. L., and Feldman, D. C. (2005). Predictors of objective and subjective career success: a meta-analysis. *Pers. Psychol.* 58, 367–408. doi: 10.1111/j.1744-6570.2005.00515.x
- Ostendorf, F., and Angleitner, A. (2004). *NEO Personality Inventory according to Costa and McCrae Revised*. Göttingen: Hogrefe.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879
- Rammstedt, B., and John, O. P. (2005). Short version of the Big Five Inventory (BFI-K): development and validation of an economic inventory for assessment of the five factors of personality. *Diagnostica* 51, 195–206. doi: 10.1026/0012-1924.51.4.195
- Sackett, P. R., and Walmsley, P. T. (2014). Which personality attributes are most important in the workplace? *Perspect. Psychol. Sci.* 9, 538–551. doi: 10.1177/1745691614543972
- Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: a commentary on four recent articles. *Perspect. Psychol. Sci.* 9, 211–218. doi: 10.1177/1745691613518074
- Schmidt, F. L., and Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychol. Bull.* 124, 262–274. doi: 10.1037/0033-2909.124.2.262
- Schmidt, F. L., and Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *J. Pers. Soc. Psychol.* 86, 162–173. doi: 10.1037/0022-3514.86.1.162
- Seibert, S. E., and Kraimer, M. L. (2001). The five-factor model of personality and career success. *J. Vocat. Behav.* 58, 1–21. doi: 10.1006/jvbe.2000.1757
- Shaffer, J. A., and Postlethwaite, B. E. (2013). The validity of conscientiousness for predicting job performance: a meta-analytic test of two hypotheses. *Int. J. Select. Assess.* 21, 183–199. doi: 10.1111/ijssa.12028
- Statistisches Bundesamt (2014). *Pay Differentials Between Men and Women 2014*. Available online at: <https://www.destatis.de/DE/ZahlenFakten/Gesamtwirtschaft/Umwelt/VerdiensteArbeitskosten/VerdiensteVerdienstunterschiede/VerdiensteVerdienstunterschiede.html>
- Stewart, G. L. (1999). Trait bandwidth and stages of job performance: assessing differential effects for conscientiousness and its subtraits. *J. Appl. Psychol.* 84, 959–968. doi: 10.1037/0021-9010.84.6.959
- Strenze, T. (2007). Intelligence and socioeconomic success: a meta-analytic review of longitudinal research. *Intelligence* 35, 401–426. doi: 10.1016/j.intell.2006.09.004
- Tabachnik, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics 5th Edn*. Boston, MA: Pearson Education.
- Woods, S. A., Lievens, F., De Fruyt, F., and Wille, B. (2013). Personality across working life: the longitudinal and reciprocal influences of personality on work. *J. Organiz. Behav.* 34, 7–25. doi: 10.1002/job.1863
- Yan, T., Curtin, R., and Jans, M. (2010). Trends in income nonresponse over two decades. *J. Offic. Stat.* 26, 145–164. Available online at: <http://www.jos.nu/Articles/abstract.asp?article=260145>
- Zinn, S., and Würbach, A. (2016). A statistical approach to address the problem of heaping in self-reported income data. *J. Appl. Stat.* 43, 682–703. doi: 10.1080/02664763.2015.1077372

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Hagmann-von Arx, Gygi, Weidmann and Grob. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**APPENDIX E: Selbständigkeitserklärung**

Hiermit erkläre ich, dass ich die Dissertation „Intelligence and its Assessment across the Life Span: Analyzing Construct and Criterion Validity“ selbständig verfasst und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe. Alle wörtlichen oder sinngemässen Zitate sind als solche gekennzeichnet. Die zur Promotion eingereichten Zeitschriftenbeiträge wurden in Zusammenarbeit mit den jeweiligen Co-Autorinnen und Co-Autoren angefertigt.

Basel, im Dezember 2016

Jasmin Gygi